STATISTICS
IN THEORY AND PRACTICE

STATISTICS
IN THEORY AND PRACTICE

BY

L. R. CONNOR, M.Sc. (Econ.), F.S.S.
BARRISTER-AT-LAW
FARR MEDALLIST IN STATISTICS

THIRD EDITION
Fifth Impression

LONDON
SIR ISAAC PITMAN & SONS, LTD.
1949

First Edition 1932
Second Edition 1934
Reprinted 1935
Third Edition 1937
Reprinted 1943
Reprinted 1946
Reprinted 1947
Reprinted 1949

SIR ISAAC PITMAN & SONS, LTD.
PITMAN HOUSE, PARKER STREET, KINGSWAY, LONDON, W.C.2
THE PITMAN PRESS, BATH
PITMAN HOUSE, LITTLE COLLINS STREET, MELBOURNE
27 BECKETTS BUILDINGS, PRESIDENT STREET, JOHANNESBURG

ASSOCIATED COMPANIES

PITMAN PUBLISHING CORPORATION
2 WEST 45TH STREET, NEW YORK
205 WEST MONROE STREET, CHICAGO

SIR ISAAC PITMAN & SONS (CANADA), LTD.
(INCORPORATING THE COMMERCIAL TEXT BOOK COMPANY)
PITMAN HOUSE, 381-383 CHURCH STREET, TORONTO

MADE IN GREAT BRITAIN AT THE PITMAN PRESS, BATH
D7—(B 2252)

PREFACE TO THE THIRD EDITION

THE new edition has been thoroughly revised and partly rewritten. New features of Part I include a section on *Seasonal Variations* and the substitution of the method of *Factorial* Moments for that of *Least Squares.* Following the Second Edition, *Small Samples* and *Analysis of Variance* have been treated in summary fashion, owing to the difficulty of doing justice to these topics in an elementary course.

Changes in Part II include the revised *Board of Trade* Index Number of Wholesale Prices, the *Actuaries'* Investment Index, the preliminary results of the Census of Production, 1935, the revised *Board of Trade* Index of Production, Clark's Quarterly Figures of the National Income, the *Economist* Index of Business Activity, etc. Other matters have been revised and brought up to date where necessary.

A novel feature is the inclusion of an Appendix on *Calculating Machines,* by Dr. L. J. Comrie, late superintendent of the Nautical Almanac Office.

Certain matter appearing in earlier editions has been excluded, either because it is now out of date, or because the information is readily available in official publications.

Attention is drawn to the *Ministry of Labour's* revised method of counting the unemployed. This was notified too late for inclusion in the final proofs.

L. R. CONNOR

PREFACE TO THE FIRST EDITION

This is the Age of Statistics. Hardly a branch of human activity does not ultimately rest upon a foundation of quantitative facts. Statistics cannot be handled effectively without a rudimentary knowledge of the laws of mass phenomena. There is no need to be mystified. Statistical methods are simple, and depend upon the consistent application of common-sense. Apparent difficulties are due to the introduction of a new order of ideas, and once the feeling of unfamiliarity is overcome, progress is rapid.

Part I provides an introduction to statistical methods for those who, from choice or necessity, handle statistical problems—the professional and business man as well as the student. Those reading for professional examinations will find it covers all, or most of their requirements, and the same remark applies to examinations in elementary statistics conducted by the universities.

Part II deals with the sources of statistical data and practical applications of method.

Elementary students are recommended to omit Chapters XII, XIII, XIV, XVII, and XVIII until they have made substantial progress.

A certain amount of elementary mathematics has been necessary, involving simple algebra up to the Binomial Theorem. Non-mathematical students may omit the algebraic proofs, together with Chapters XVII and XVIII. The said proofs are not essential to the understanding of the text; but if omitted, a certain amount must be taken on trust.

Acknowledgments are due to H.M Stationery Office, the Royal Statistical Society, the London and Cambridge Economic Service, the Bank of England, and the Society of Incorporated Accountants and Auditors, for permission to reproduce copyright matter. Further acknowledgments are made in the text

Thanks are due to numerous friends who have read the proofs in whole or part. Their criticisms and suggestions have proved invaluable.

L. R. CONNOR

CONTENTS

PART I

STATISTICAL METHOD

PART II
APPLIED STATISTICS

LIST OF TABLES

LIST OF FIGURES

STATISTICS
IN THEORY AND PRACTICE

CHAPTER I

INTRODUCTION

Scope and Definition.

Statistics (Plural) are measurements, enumerations or estimates of natural or social phenomena, systematically arranged so as to exhibit their inter-relations. It is implied that the phenomena in question are subject to uncontrolled variations (due to the workings of complex causal systems), which impair unduly the significance of individual data, but to a less extent the significance of data taken in bulk.

Statistics (Singular) is the branch of applied mathematics which specializes in data of this kind.

An Exact Law (characteristic of the physical sciences), when correctly formulated, is held to be true of every individual case coming under its jurisdiction, whereas a Statistical Law is only held to be true on the average or in the long run. Broadly speaking, exact laws reign in the physical sciences and statistical laws elsewhere. The distinction however is largely one of degree. Some so-called "exact" laws are recognized to be statistical laws in a disguised form, e.g. the kinetic theory of gases. According to some modern schools of thought, "exact" laws do not exist, the whole of the Universe being ultimately statistical.

The practical effect is that statistical data must always be treated as approximations or estimates and not as precise measurements. If the data are statistically uniform, i.e. if they belong to a causal system which does not change its main characteristics over the field of observation, it is possible to make critical tests of the significance of the results by applying simple theorems in the calculus of probability, but if they are not statistically uniform, the critical methods break down and we are forced to rely on broad judgments which are not precise.

Since the laws of Nature do not change in our experience, it is reasonable to suppose that critical tests will succeed in the biological and allied sciences, the more especially when experimental methods are available. In the social sciences, experiments are not usually possible, and the statistical background is continually affected by the changeability of human institutions, so that there is a presumption that critical methods will not succeed except when applied on a large scale to data of a highly stable type. Neglect of this distinction has been the cause of much bad work and has tended to bring the science into discredit.

In order to develop elementary statistical methods in clear and intelligible fashion, it is necessary in the first instance to ignore the complications due to want of statistical uniformity. This amounts to postulating that we are working with ideal data. It must not be supposed, however, that a method put forward as an illustration of a particular proposition may be applied indiscriminately, in any circumstances. As in other sciences, it is necessary to begin with clear statements of principle, applied to simplified versions of the facts, and to lead the student by degrees to the numerous qualifications required to apply his knowledge to the more complex conditions which obtain in practical problems.

Main Divisions of Statistics.

The two main divisions of statistical science are—

1. Statistical Method, dealing with general rules and principles common to all branches of data.

2. Applied Statistics, dealing with the application of these methods to concrete subject-matter.

The following classification of *Applied Statistics* is based upon practice and expediency, and does not pretend to be exhaustive nor perfectly logical—

1. Biometry, dealing with measurement of living organisms.

2 Psychometry, dealing with mental phenomena.

3 Vital Statistics and Demography, which study population movements

4. Administrative Statistics, collected for governmental purposes.

5. Social Statistics, dealing with measurement of social phenomena.

6. Economic Statistics is properly a subdivision of *Social*

Statistics. Owing, however, to peculiarities of technique, it is more convenient to regard it as a separate branch.

7. Business Statistics involves special problems and may also be regarded as a separate branch.

The present work deals with the subject mainly from the standpoints of divisions (4) to (7), but illustrations will be drawn from other sources where necessary.

CHAPTER II

ORGANIZATION OF A STATISTICAL INQUIRY

At the risk of labouring the obvious, it is necessary to insist upon a careful preliminary study of the inquiry to be undertaken, for more effort is wasted by inattention to this maxim than from any other cause. Not only must one consider *what* information is required, but *by whom* it is required. Modes of treatment will vary according to the purposes for which the data are collected: whether for private and confidential use or for publication in the general interest. Trade associations collect confidential information from members for their own benefit; some of the figures are passed on in bulk to the *Board of Trade* for public information. Official, scientific, and commercial inquiries will visualize the same subject-matter from different angles; the needs of the specialist are not identical with those of the non-specialist, and facts and distinctions relevant to one class of investigation will be immaterial to another.

Points of an Inquiry.

The following points deserve special attention—

1. Is the desired information capable of statistical treatment? If not, is there some allied class of information reducible to numerical standards?

Intelligence cannot itself be measured directly. Therefore, reliance must be placed upon some external index, such as marks gained at an examination or intelligence test. Crime involves not only action but also intent. In practice the sociologist must be satisfied with statistics of criminal convictions.

2. What is the precise nature of the object to be measured?

An inquiry into a wage problem would involve consideration of the following points—

(*a*) Should the inquiry relate to wage rates or to actual earnings?

(*b*) If to wage rates, should it refer to rates recognized by trade associations or to rates actually paid? Should allowances be made for overtime, undertime, and bonuses?

(*c*) Should receipts in kind be included, and allowances made for special expenses?

(*d*) Should supervisory and clerical workers be included?

3. What should be the field of inquiry?

The *Census of Production* includes extractive and manufacturing industries. Transport, distribution, and personal services are excluded because of the difficulty of finding suitable units of output. The Census also omits small workshops on the ground that the details obtainable are not worth the trouble of collection.

4. Should the inquiry be comprehensive or by sample?

The *Population Census* is comprehensive because an exact enumeration is essential for administrative purposes. The *Ministry of Labour's* periodical inquiries into the circumstances of applicants for unemployment benefit are conducted by sample, because the trouble and expense of a comprehensive inquiry would be disproportionate to the extra reliability achieved.

5. What co-ordination and standardization problems are involved?

The headings of the *Census of Production* schedules are chosen to agree as far as possible with the headings of the Import and Export list, so as to facilitate comparison between production and foreign trade. On the other hand, the classification of industries employed in the *Production Census* differs from that employed in the corresponding *Wage Census*. Consequently, only broad comparisons between production and wages are possible.

6. What is the degree of accuracy (*a*) ideally demanded, (*b*) actually attainable?

It is implied that statistics shall be compiled according to reasonable standards of accuracy, but the latter vary according to the conditions of the problem. Much depends upon the purpose of the inquiry, the type of informant, and the time and labour involved in checking up his statements. Efforts spent in securing representativeness and freedom from bias are more remunerative than efforts spent on needless detail.

The *Population Census* aims at accuracy to the last person, and actually achieves it to within a few thousand. Whilst a less degree of accuracy would be sufficient for administrative purposes, it is important to guard against the danger of any considerable body of persons being omitted, and this purpose is best secured by compelling everyone to give an account of himself. The *Ministry of*

Agriculture's Crop estimates are the result of reasoned guesses by skilled observers. Precision is inessential, but speed is imperative.

Continuous or Repetitive Inquiries.

If the inquiry is a continuation or repetition of previous inquiries, it is merely a question of following the old plan, subject to such minor amendments as experience may suggest. Since, however, continuity of information is a major objective, substantial changes in the form or contents of the tables are to be deprecated. At intervals, however, the whole problem must be studied afresh. Conditions may have changed, certain kinds of information become obsolete, and new opportunities and requirements materialized. In these circumstances advantages of improvement must be weighed against advantages of continuity. In any event a complete break should be avoided. It is frequently possible to arrange for data to be collected and tabulated on both systems—the old and the new—for a short period, with a view to easing over the transition. Numerous examples of changes and expedients to overcome them may be found in the *Statistical Abstract for the United Kingdom.*

CHAPTER III

STATISTICAL DATA

Statistical Data are usually classified as Primary and Secondary. Primary consist of the raw material of inquiry, whilst secondary consist of material that has been worked up to some extent. This distinction is not, however, clear-cut, for secondary data at one stage may become primary data of the next.

Collection of Material.

There are several ways of collecting material for statistical inquiries.

Primary Data.

1. *Direct Personal Observation* is not usually feasible, but may be employed in laboratory experiments and localized inquiries.

2. *By Personal Interviews*, assisted by a standard list of questions. This method is useful when the information desired is complex or there is reluctance or indifference on the part of informants.

3. *By schedules distributed and collected by enumerators*, who assist the informants, where necessary, to fill them in. This is the best plan to follow when a large-scale inquiry is in progress.

4. *By forms, schedules, or questionnaires sent and returned through the post.* This plan is cheap and fairly expeditious, provided the informants can be relied upon to answer intelligibly. It is, of course, the standard method for routine business and administrative inquiries.

5. *By estimates from local sources.* Under this plan there is no formal collection of data, but local agents or correspondents are asked to send in estimates, using their own judgment as to the best way of obtaining them. This method is useful when figures are required cheaply and expeditiously, and accuracy is not of prime importance.

Secondary Data.

6. *By utilizing information collected by other agencies or for other purposes.* Income-tax returns were used on behalf of the *Colwyn*

Committee to prepare estimates of commercial profits; and forms completed by parents of scholarship holders have been used in connection with inquiries into social conditions of the working classes.

7. *By utilizing published sources of information*, e.g.—

(*a*) *Official*, published by Government Departments, Municipalities, Public utility undertakings, Royal Commissions, etc.

(*b*) *Trade Association* statistics and the like.

(*c*) *Technical and trade journals.*

(*d*) *Research agencies*, such as the Universities, the *London and Cambridge Economic Service*, etc.

8. *Business services and agencies*, such as the *Moody-Economist Service.*

Editing Primary Data.

Primary material should be scrutinized at an early stage with a view to the detection of errors, omissions, and inconsistencies. If possible, defective schedules should be returned for amendment, but there is no objection in principle to the investigator correcting them himself provided he has reasonable grounds upon which to act. Thoroughly unsatisfactory schedules must be rejected.

Sometimes the nature of the answers indicates that a question has been badly drafted, or again the answers may bring new ideas to light or suggest the need for information not previously desired. In such cases, it is a nice question whether to let the information stand, to correct it oneself, or to send out supplementary schedules Experience is the best guide.

Editing Secondary Data.

Secondary data should never be accepted without careful inquiry and criticism In particular, the investigator should satisfy himself as to—

1. The standing and reliability of the compilers;
2. The scope and object of the inquiry;
3. The sources of the information;
4. The degree of accuracy aimed at and achieved.

Statistics, especially other people's statistics, are full of pitfalls for the user. Terms may be used in peculiar senses; meanings may

have imperceptibly changed; external factors may operate to produce discrepancies. Typical sources of difficulty are as follows—

1. *Changes in form and content over a period of years.* Deliberate changes are usually specified, but involuntary changes due to readjustment of ideas or transfers of personnel are more elusive. It is a safe rule that statistics are seldom truly comparable over a long period of years.

2. *Changes due to variations in definition.* The significance of the income-tax statistics is altered every time the exemption limit is changed.

3. *Changes in geographical and administrative areas.* A recent instance is provided by the exclusion of the *Irish Free State* from *United Kingdom* statistics. Readjustments of local government boundaries furnish a perennial source of difficulty.

4. *Inadequacy or incompleteness of the information provided.* The risk of error is the greater when defects are not obvious. The *Ministry of Labour's* unemployment figures relate only to insured persons. Foreign trade statistics relate only to tangible goods, not to services. The *Board of Education's* statistics only refer to State-aided or State-inspected education.

Specimen Questionnaire.

The following is a specimen questionnaire suitable for an inquiry into the cost of living for middle-class families—

INQUIRY INTO COST OF LIVING

Code No...............

SECTION I

Surname....................	Profession or Occupation................................
Christian Names	State whether—
Address	Employer or Managerial ....
....	Employee ....
State whether occupying—	Independent Worker....................
A separate dwelling	
Unfurnished rooms	
Furnished rooms without service	
Furnished rooms with service	Industry ....

SECTION II

Constitution of Household

	Male	Female
Householder(s)		
Wife		
Dependent children—		
Aged 0–5		
,, 5–16		
,, 16 and over		
Other dependants		
Servants (full-time)		
Children (not dependent)		
Other persons		
Totals . . .		

SECTION III

Family Income

	Year 1936	
Family Income—	£	£
Head of Household .		
Main occupation . . .		
Supplementary earnings . .		
Wife's earned income . . .		
Other earned income		
Income from investments		
Income from house property, etc . :		
Other items		
Contributions towards household expenses		
Children		
Other persons		
Abnormal or Non-recurrent items (specify) .		
Total . . .	. . £	

SECTION IV

Family Expenditure

	Year 1936	
	£	£
Family expenditure—		
Rent (or annual value of house) . .		
Rates and taxes		
Housekeeping		
Heat, light, and water		
Repairs and renewals		
Clothing		
Education		
Medical attendance		
Holidays and recreation . . .		
Insurance		
Motor car		
Other items		
Abnormal or non-recurrent items (specify)		
Total . . . . . . £		

SECTION V

Surplus or Deficit

	£	£
Surplus (state how disposed of) . . .		
	£	

	£	£
Deficit (state how met)		
	£	

Remarks ....
....
....
....

Notes

1. Income should be given net, but this does not apply to income-tax.
2. The figures should relate to the year 1936. If actual figures are not available, estimates should be given.
3. Abnormal or non-recurrent items should be distinguished.

and statistics of railway movements are accordingly expressed in terms of ton-miles.

Even the composite unit may not avoid the difficulty altogether. Jobs differ in intensity of application required, and men differ in strength and skill. Ton-miles differ in quality according to ruling gradients and curves. On the other hand, there are limits to which distinction may be practicable, and disturbances of this class tend to disappear when numbers are large.

Statistical Variates.

Any set of facts whatever may form the subject of statistical measurement, provided the individual facts may be specified and identified. The essence of the process consists in (1) classifying or grading the various kinds of facts according to their relevant characteristics (qualities or attributes) and (2) counting or measuring the items in each class or grade. The set of characteristics is known collectively as the Variate and the results of counting or measuring as the *Frequency* or *Measurement* respectively. The Variate may be either quantitative or qualitative and in the former case either continuous or discrete.

A variate is said to be quantitative when it can be measured numerically (e.g. age, height, income, examination marks), and qualitative when it can only be described verbally (e.g. hair colour, occupation, geographical location). A variate is said to be continuous when it may pass from one value to the next by indefinitely small gradations (e.g. height) and discrete when there are gaps between one value and the next (e.g. children per family). The distinction between qualitative and quantitative variates is statistically important, since the latter lend themselves to mathematical treatment in the shape of averages, etc., whereas the former do not. The other distinction is less important, and for the most part, the elementary student can afford to ignore it. An important case arises when the variate represents succession in Time. Instead of studying a set of different facts under the same conditions we have to study the same set of facts under different conditions. This involves special problems, which must be elucidated.

Statistics and Causation.

We distinguish between assignable causes (traceable to a specific

origin) and chance causes (combinations of small factors that cannot be individually identified).

Any quantity varying in ways that can be accounted for by the operation of a constant system of chance causes is said to be statistically uniform. It is one of the objects of analysis to break up complex aggregates into groups complying with this condition.

CHAPTER V

CLASSIFICATION AND TABULATION

I. Classification

Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities, and gives expression to the unity of attributes that may subsist amongst a diversity of individuals. Its objects are to display points of similarity and dissimilarity, to save mental effort by the systematic suppression of irrelevant detail, to enable one to form mental pictures of objects of perception or conception, and to suggest bases for comparison and inference.

Classification is an organic process of thought, implicit in every judgment, however crude. Classifications of ordinary thought and speech are mainly utilitarian, having been worked out by a long process of trial and error, and are adjusted to the mentality of the average man in his everyday dealings.

Scientific classification involves the same processes of thought, but here they go deeper, and aim at a logical arrangement of things according to their fundamental properties and characteristics.

An ideal classification would imply the arrangement of the entire contents of the universe in mutually exhaustive categories. Everything would have its place, and there would be no room for doubt or ambiguity. Such a classification would be germane to a repository of universal knowledge, but far too cumbersome for practical scientific use. Properties highly significant in one branch of inquiry may be ignored in another. Classification is therefore many-sided: each department of knowledge must possess its own scheme, stable enough to facilitate the ready exchange of current thought, yet flexible enough to permit the absorption of new ideas as they materialize.

In statistics, classification is a preliminary to enumeration, for the subject-matter of statistical science must be capable, not only of characterization, but also of measurement.

Statistical Classification.

Statisticians usually adopt as basis the classification current in the branch of inquiry with which they are dealing; but this does not preclude them from merging distinctions that are statistically irrelevant and subdividing categories that are statistically significant.

Distinctions of hair and eye colour interest the ethnologist, but are irrelevant for most other purposes. Coins, bank-notes, and cheques are different varieties of monetary instruments with different characteristics, but since they all possess the common property of settling debts, they may be merged under the single term "currency."

Classification *simpliciter* involves the construction of a system of labelled compartments into which individual items may be thrust as and when they come to notice, but does not attempt to account for every item that might possibly occur.

On the other hand, statistical science aims at accounting for every item. By distributing a definite population or sample from that population among the compartments so that nothing is left over, it places the operation upon a numerical basis and assigns to each class its proper importance in the scheme of things.

Bases of Classification.

From the standpoint of Measurement[1] we distinguished variates as quantitative and qualitative. From our present standpoint it is necessary to refine upon this distinction.

Logically speaking, classification may be performed upon either of four distinct bases—

1. Quantitative, when the basis of distinction rests upon differences in quantity. An analysis of sales according to differences in the weight, volume, or value of the goods involved in each transaction would be quantitative.

2. Temporal, involving the time at which the objects in question were measured or the events in question occurred. An analysis of annual sales by weeks, months, or quarters involves temporal classification.

3. Spatial, referring to the distribution of items in space, e.g. annual sales by geographical areas.

4. Qualitative, in the narrower sense when the basis of distinction

[1] Chapter IV, p. 14.

rests upon differences in quality or condition. An analysis of sales by reference to the kind of goods sold involves qualitative distinctions.

Seriation.

Closely allied with classification is the process of Seriation. If two variable quantities can be arranged side by side so that measurable differences in the one correspond with measurable differences in the other, the result is said to form a statistical series.

An enumeration of the population of England and Wales at successive dates would form a statistical series, because one measurable quantity (population) is tabulated against another measurable quantity (time). In the same way an analysis of the population by ages would form a serial distribution because here again, a measurable quantity (population) is associated with another measurable quantity (age). An analysis of the population by places of residence or by occupations would, however, not form a series, because the basis of distinction (residence in the one case and occupation in the other) is not measurable.

Examples of Classification.

Table 1 (p. 19) involves classification by time, space, and quality.

Table 2 (p. 19) involves classification by time, space, and quantity.

Fineness of Classification.

The question of the fineness of the classification to be adopted now demands attention. The essential feature of classification is that for the purpose in hand all objects comprised in the class or group are treated as similar. The finer the network is drawn, the less the chances of including unlike objects under the same head, but the more numerous the headings become, the more complicated and unmanageable the table. Moreover, a fine classification tends to produce irregularities in the figures, since the smaller the number of objects included under one head, the less is the chance of abnormalities cancelling out. On the other hand, too coarse a classification tends to obscure essential facts and differences. The larger the number of objects involved, the finer may the classification be. It is often necessary to experiment with different systems, adopting that which appears to secure the greatest balance of advantage.

TABLE 1

VESSELS OF 100 TONS GROSS AND UPWARDS ON THE REGISTERS OF PRINCIPAL COUNTRIES AT MID-YEAR, 1913, 1924, AND 1930.[1]

COUNTRIES	ENTERED IN THE REGISTER FOR 1913		1924			1930		
	Steam and Motor	Sailing	Steam	Motor	Sailing	Steam	Motor	Sailing
	(Steam and motor ships in thousand tons gross. Sailing ships in thousand tons net.)							
Great Britain and Ireland . .	18,274	422	18,427	527	152	18,060	2,262	117
Australia and New Zealand	1,575	160	789	10	10	626	52	7
Canada .			1,069	11	109	1,132	103	97
Other Dominions			613	20	62	831	45	51
TOTALS .	43,079	3,891	59,538	1,976	2,510	59,928	8,096	1,584
	46,970		64,024			69,608		

[1] Figures from *Lloyd's Register of Shipping* as quoted in *Statistical Tables Relating to British and Foreign Trade and Industry* (1924–1930) (Cmd. 3737), Part I, p. 218

TABLE 2

TONNAGE CLASSIFICATION OF STEAM AND MOTOR VESSELS BY NATIONALITY AT MID-YEAR, 1924 AND 1930[1]

COUNTRIES	On Register for	100 tons gross and under 2,000 tons	2,000 tons gross and under 4,000 tons	4,000 tons gross and under 8,000 tons	8,000 tons gross and under 10,000 tons	10,000 tons gross and under 20,000 tons	20,000 tons gross and over
Great Britain and Ireland	1924	2,573	2,847	9,202	1,572	2,162	598
	1930	2,434	2,305	10,433	1,648	2,645	857
Rest of British Empire .	1924	817	740	792	92	151	—
	1930	950	684	892	104	158	—

[1] Figures from *Lloyd's Register of Shipping*, as *loc. cit.*, p. 219.

II. Tabulation

Tabulation involves the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration.

The original material must be dissected and re-arranged before it can be utilized for statistical purposes.

Stages in Tabulation.

In typical cases, involving the analysis of a number of schedules containing information on a variety of points, there would be at least three stages in the operation, viz.—

1. Extracting each item of information from the schedules and listing the results on working sheets ruled with appropriate headings.

2. Summarizing the working sheets and transferring the results to a fresh set of sheets which form the subject of actual study and experiment, and the basis for any further manipulation that may be necessary.

3. Preparing the final tables embodying the results of the inquiry *in so much detail as may be necessary. At this stage all experimental* figures are eliminated as well as other intermediary figures which, although needed for purposes of calculation, are not required for the understanding of the final result.

Original and Derivative Tables.

Statistical tables may be either original or derivative. Original tables present the information in substantially the same form as it was collected, whilst derivative tables imply some process of manipulation, such as grouping, totalling, averaging, or other operations of a mathematical nature.

Working Sheets.

No general rules can be laid down for the drafting of working sheets, but the following hints may be useful—

1. Allow plenty of room for amendment and correction.

2. Do not attempt too much at once. Analyse the material stage by stage, arranging for a control column as a check upon the results. Check and cross-cast at each stage before the next is begun.

3. Give the sheets proper titles and the columns their proper headings. Number the sheets consecutively as a precaution against loss.

4. When it is a question of counting numerous items, the following device saves time and space—

Put a stroke for every item counted, as follows—

/	//	///	////	~~////~~
1st	2nd	3rd	4th	5th item

The groups of five stand out, and are easy to count.

Other Hints on Tabulation.

Although tabulation is a routine process, a careful watch must be kept at all stages in order to prevent misunderstandings and to secure uniform application of principle throughout the operation. Columns that show a tendency towards overcrowding should be broken up, whilst other columns may need merging. Sometimes the draft rulings are found to be unsuitable, and the process has to be recommenced. It is advisable to insert control columns, so that arithmetical accuracy may be secured by a system of cross-casting. When the working sheets are completed, totalled, and checked, results are inspected and tested, and it is decided whether to proceed upon the lines originally chosen or to experiment with some new combination. This experimental stage may require considerable time, and much work may have to be discarded before the inquiry is fit to proceed.

Card Systems.

When a complicated system of cross-classification is in view, it may be preferable to transfer the original data to separate cards, which can be sorted and re-sorted in any manner desired. Cards of various colours may be used for different classes of data, or the same effect may be obtained by means of notches in the margins located at various points.

Mechanical Tabulation.

The subject of Mechanical Tabulation is dealt with in Appendix III, contributed by Dr. L. J. Comrie.

Construction of Tables.

The basis of a statistical table is a column or row of compartments or cells, each bearing a label indicative of its contents. The

number or measurement of the items with the characteristics implied by the cell label is inserted in the cell. This is simple tabulation.

Tabulation in two dimensions involves both columns and rows of cells in combination, so that a given cell may be referred to two headings at one and the same time.

Tabulation in three or more dimensions is not possible in the literal sense. But the same effect may be obtained by subdivision of rows or columns, or by breaking up the table into sections.

SPECIMEN TABULATIONS

Simple Tabulation.

X COMPANY—TRADING PROFIT

Year (1)	Total Trading Profit (2)
	£

The above is a case of simple seriation, the dependent variable (Total Trading Profit) being tabulated against the independent variable (Time).

In the Table on p. 23 the entries have been analysed into their constituents by means of a scheme involving a combination of co-ordinate and subordinate classification. Columns (5), (6), and (7) are subordinate to column (8), columns (8) and (9) to column (4), and columns (3) and (4) to column (2).

The table on p. 24 involves a further subordinate cross-classification according to type of product.

The table given on p. 25 involves a still further element of subordinate classification under branches, but in order to make the table manageable, it has been necessary to abridge the analysis of profit.

It also represents the apparent limit to which cross-classification can be carried in this instance. Any further subdivision must be affected by breaking the table up into separate sections. The most promising developments appear to lie in the direction of separate tabulation either under years or under products.

Twofold Tabulation.

X Company—Trading Profit

Year	Gross Sales	*Less* Selling Expenses	Net Sales	*Less* Costs				Balance (= Trading Profit)
				Material	Labour	Overheads	Total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	£	£	£	£	£	£	£	£

Threefold Tabulation.

X Company—Trading Profit

Year	Product	Gross Sales	*Less* Selling Expenses	Net Sales	*Less* Costs				Balance (= Trading Profit)
					Material	Labour	Overheads	Total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		£	£	£	£	£	£	£	£
1935	A								
	B								
	.								
	.								
	Total								
1936	A								
	B								
	.								
	.								
	.								
	Total								

Fourfold Tabulation.

X COMPANY—TRADING PROFIT

Year	Product	Net Sales				*Less* Costs				Balance (= Trading Profit)			
		Branch P	Branch Q	Branch R	Total	Branch P	Branch Q	Branch R	Total	Branch P	Branch Q	Branch R	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
		£	£	£	£	£	£	£	£	£	£	£	£
1935	A												
	B												
	. . .												
	Total .												
1936	A												
	B												
	. . .												
	Total .												

The forms are intended to be only illustrative. There are, of course, numerous other ways of arranging the same matter.

Rules for Construction of Tables.

The following rules are intended as a guide to the construction of Statistical Tables, but are not intended to be exhaustive.

1. First make out a rough draft.

2. Avoid complicated tables. Information of a high degree of complexity should be broken up into sections. Introduce summary tables where they are likely to be useful.

3. See that the title is specific and comprehensive. Be careful about geographical areas and dates. Always give the source of information. Some official figures relate to *England and Wales,* others to *Great Britain*, and others to the *United Kingdom.* Precision on such points is essential.

4. Consider the column and row headings carefully, and make sure they express exactly what is intended. Remember to state the unit of measurement. Insert definitions and explanatory notes where necessary, and draw attention to any important changes of definition or content that may have occurred between one year and another.

5. Consider the contents of the cells. Some may be overcrowded and need breaking up, and others may need merging. Avoid blank spaces as far as possible.

6. With a large or complicated table, the columns or rows should be numbered or lettered for identification purposes. If the table is very wide it is useful to repeat the row headings on the right-hand side.

7. With a long table it is advisable to leave a break at every fifth line.

8. Derivative figures, e.g. averages or percentages, should be placed as close to their originals as possible.

9. Study the possibilities of distinctive type. Heavy type may be used for totals or important figures. Italic type is available for several purposes, e.g. minus quantities; last year's figures as compared with this year's; ratios, percentages, or averages; estimates as contrasted with actual figures.

10. The possibilities of light and heavy rulings should be considered.

11. Totals are usually placed at the foot of the table, but there is much to be said for the growing practice of placing them at the head.

12. The cost of printing tabular matter is high, and it may be necessary to break some of these rules to save expense.

13. Consider the degree of accuracy necessary. For most practical purposes three- or four-figure accuracy is sufficient. Much labour and space can be saved by tabulating to the nearest hundred, thousand, or million

CHAPTER VI

DIAGRAMS

CURRENT usage does not distinguish clearly between the terms *diagram, graph, chart,* and *figure.* For convenience' sake it is proposed to employ the word *diagram* with reference to geometrical constructions such as parallelograms and circles, expressive of simple statistical relationships, and to reserve the word *graph* for more complicated constructions involving continuous curves. The words *chart* and *figure* can be used indifferently of both

Basis of the Diagram.

Since a statistical quantity is usually a concrete number, expressing the results of counting or measurement, or an abstract number such as an average or percentage, it may be represented by a linear magnitude drawn according to a pre-arranged scale. Lines do not, however, stand out well, and in practice it is usual to substitute bars of uniform width. (See Figs. 1 and 2.)

Comparisons involving different quantities or the same quantity at different dates may be effected by drawing two or more bars in proximity.

The actual figures may be inserted at the side as shown.

Division of a Whole into Parts.

The bar diagram may also be used to exhibit the division of a whole into its component parts. (See Fig. 3)

Fig. 4 introduces new features, viz.—

1. The statistics relate to abstract quantities (proceeds, costs, and profits per ton).

2. Negative quantities (losses) are disposed of by the device shown.

3 The diagram is arranged vertically instead of horizontally.

Percentage Distributions.

The distribution of an aggregate into its parts may also be effected upon a percentage basis, and in that case there is the option of

EXPORTS OF U.K. PRODUCE & MANUFACTURES

BY COUNTRIES OF CONSIGNMENT: 1929 & 1932-6

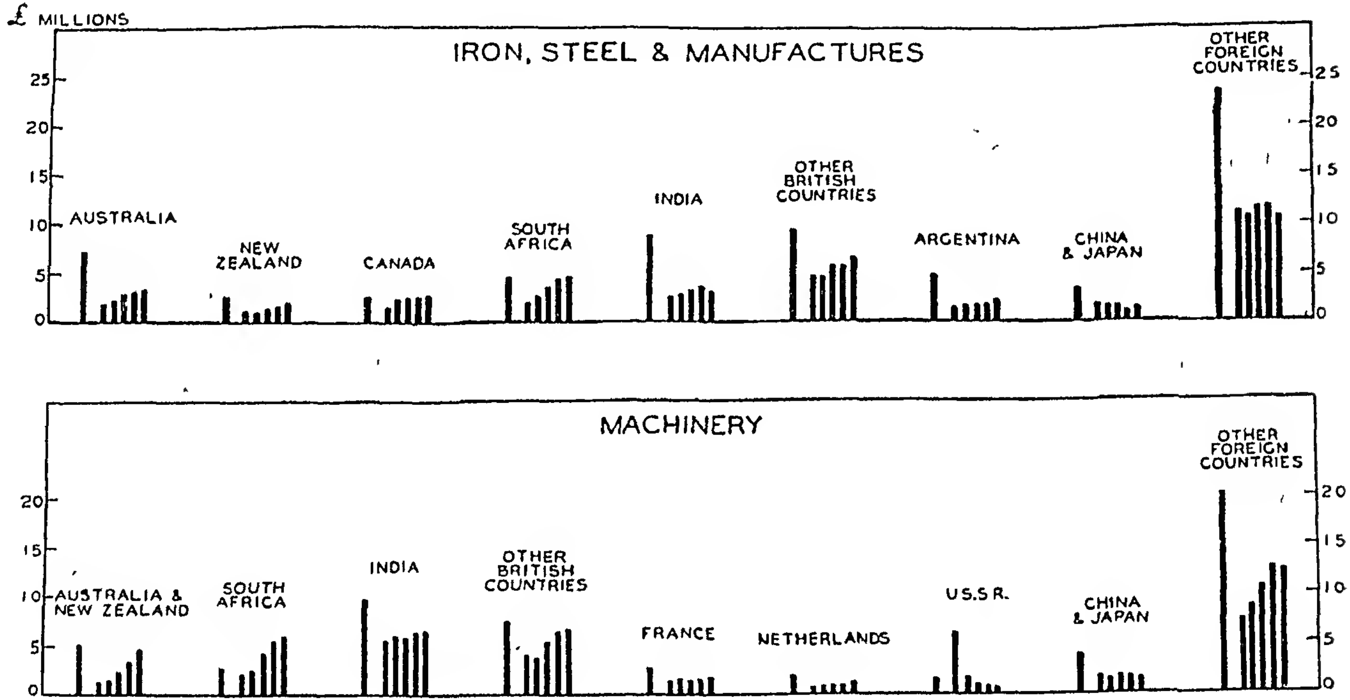


FIG. 1

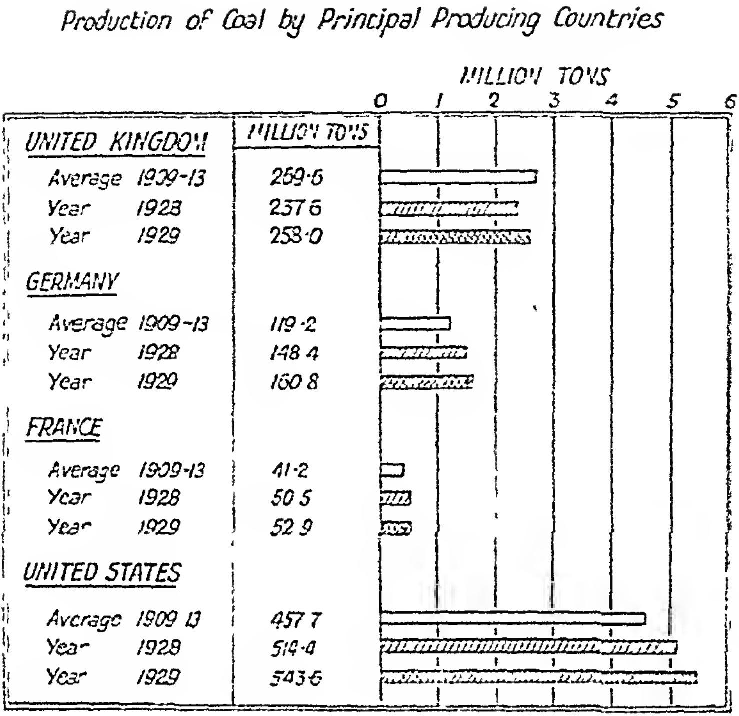

Production of Coal by Principal Producing Countries
MILLION TONS
0
1
2
3
4
5
6
MILLION TONS
UNITED KINGDOM
Average 1909-13 259·6
Year 1928 237 6
Year 1929 258·0
GERMANY
Average 1909-13 119·2
Year 1928 148 4
Year 1929 160 8
FRANCE
Average 1909-13 41·2
Year 1928 50 5
Year 1929 52 9
UNITED STATES
Average 1909 13 457 7
Year 1928 519·4
Year 1929 543·6

FIG 2

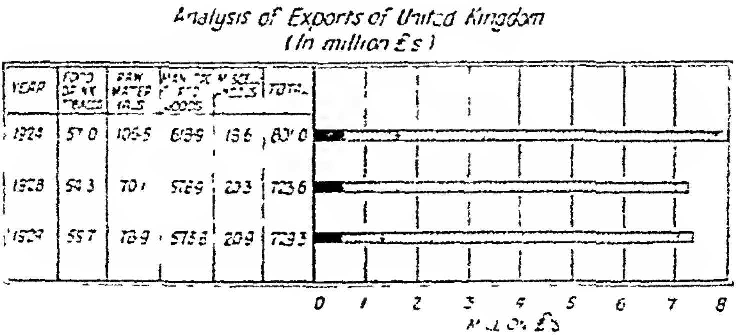

Analysis of Exports of United Kingdom
(In million £s)
YEAR
TOTAL
0
1
2
3
4
5
6
7
8

FIG. 3

TABLE 3

PRODUCTION OF COAL IN DURHAM DISTRICT—PROCEEDS, COSTS, AND PROFITS OR LOSSES PER TON,[1] 1924 and 1928

	1924	1928
	s.	s.
Proceeds per ton disposable commercially . .	19·91	12·16
Costs per ton disposable commercially—		
Wages	12·76	7·95
Other costs	5·46	4·51
Royalties	0·54	0·50
	18·76	12·96
Profit (*Loss*) per ton . . .	1·15	*0·80*

[1] *Statistical Tables Relating to British Foreign Trade and Industry*, 1924–30 Cmd. 3849), Part II, p 13.

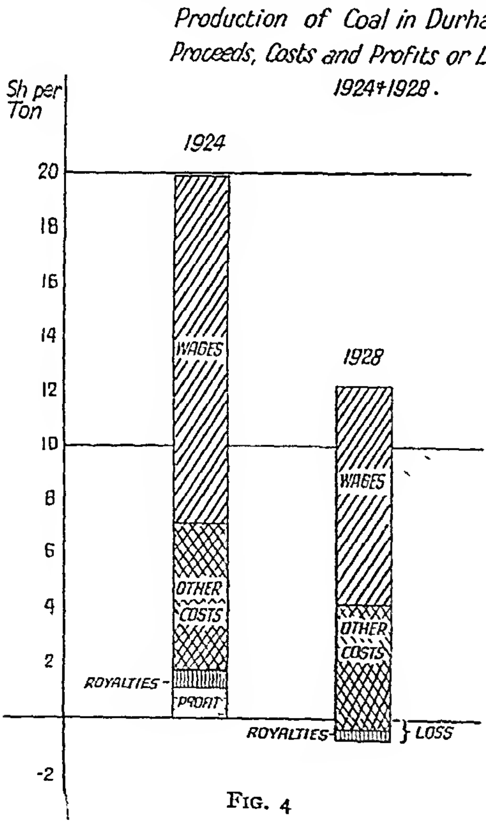


FIG. 4

presenting the results in the form either of a bar or a "pie" diagram.

For the bar diagram the figures must be recalculated upon a percentage basis, and for the "pie" diagram upon the basis of the 360 degrees in the circumference of a circle.

The following table gives the necessary calculations in respect of the coal production figures of Table 3—

TABLE 4

PRODUCTION OF COAL IN DURHAM DISTRICT—PROCEEDS, COSTS, AND PROFITS OR LOSSES PER TON, 1924 and 1928

PROPORTIONATE BASIS

	1924			1928		
	s	%	Deg	s	%	Deg
Proceeds per ton disposable commercially . . .	19.91	100.0	360.0	12.16	100.0	360.0
Costs per ton disposable commercially—						
Wages	12.76	64.1	230.7	7.95	65.4	235.4
Other costs	5.16	27.4	98.7	4.51	37.1	133.5
Royalties	0.51	2.7	9.8	0.50	4.1	14.8
	18.76	94.2	339.2	12.96	106.6	383.7
Profit (*Loss*) per ton	1.15	5.8	20.8	*0.80*	*6.6*	*23.7*

Colours provide a useful aid to the interpretation of statistical diagrams, but their use in this volume is prohibited by considerations of expense.

Pictures and Maps.

Statistical facts are sometimes represented by pictures of various sizes. Since, however, pictures introduce the element of area, or of volume, which the eye finds difficult to measure and interpret, they are not recommended for serious statistical work.

Spatial Distributions may often be represented by Statistical Maps. Dots may be used for isolated items, and where items are numerous, their approximate density may be indicated by means of shading.

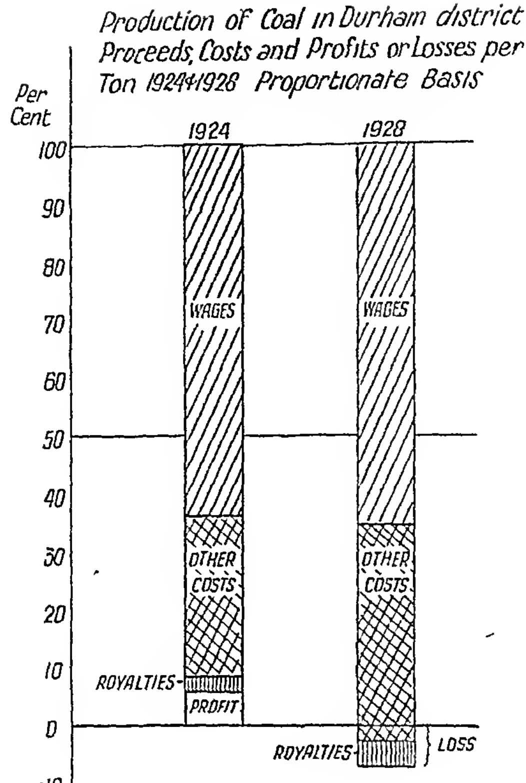


FIG. 5

Production of Coal in Durham District. Proceeds, Costs and Profits or Losses per Ton. Pie Diagrams.

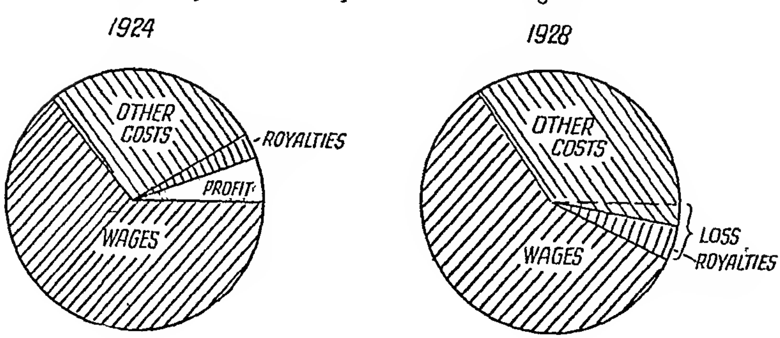


FIG. 6

The following map,[1] showing the number of acres of arable land per 100 acres of crops and grass in 1925, is reproduced by permission of H M. Stationery Office.

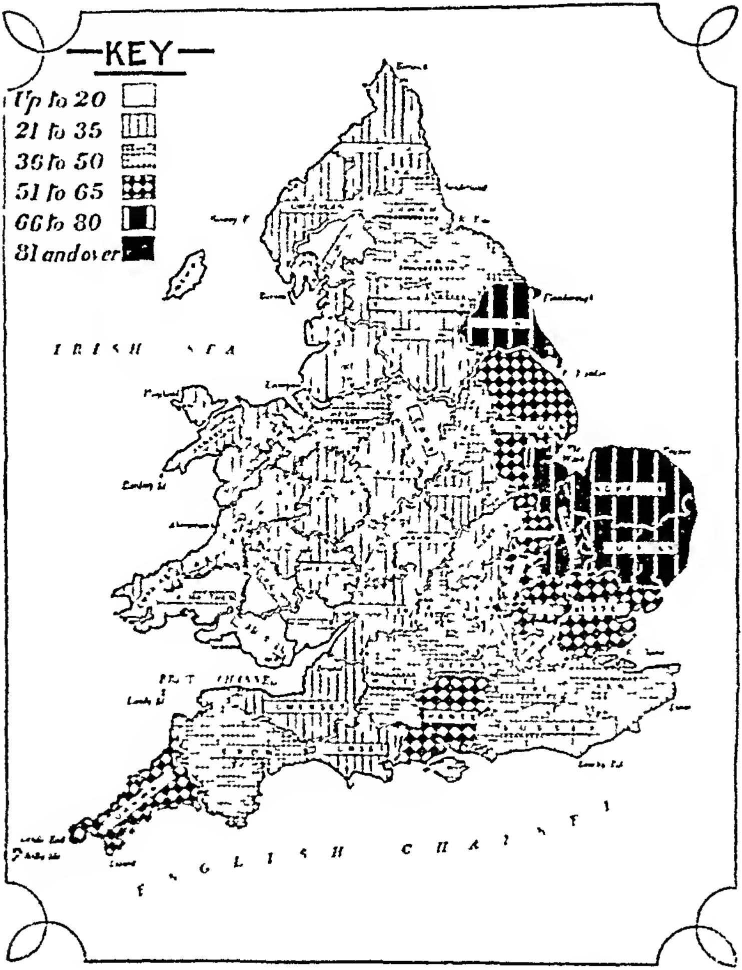


Fig. 7 Number of Acres of Arable Land per 100 Acres of Crops and Grass in 1925

[1] *The Agricultural Output of England and Wales, 1925*

Another method involves the use of a large-scale map fixed to the wall, with the information indicated by means of discs of coloured paper or pins with coloured heads. This method is particularly useful when it is required to keep information continuously up to date.

Further examples are precluded by lack of space. Those interested should consult specialized works on the subject.[1]

[1] The best work is Karsten's *Charts and Graphs*.

CHAPTER VII

GRAPHS

Regular and continuous series of data lend themsevles to treatment by graphic methods involving continuous curves.[1] Graphic methods are more powerful than diagrammatic methods, for not only do they effectively illustrate the facts presented, but they suggest new relations that may not become apparent from a study of the figures themselves. On the other hand, it is easy to draw graphs that are misleading.

Rectangular Co-ordinates.

The most useful variety of graph is that represented by rectangular co-ordinates in two dimensions. The framework of the system consists of two straight lines intersecting at right angles. The horizontal line is the x-axis or axis of abscissae, and the vertical line the y axis or axis of ordinates. The origin or zero point of the graph is located at the intersection of the two axes. Distances measured towards the right or upwards are reckoned as positive, and distances measured towards the left or downwards as negative. Any point in the four quadrants into which the graph is divided may be located unequivocally by reference to two co-ordinates drawn parallel to the axes of reference.

The scales of measurement may be chosen at convenience, and there is no necessary connection between the x and the y scales.

In Fig. 8 the scale of the x axis is twice that of the y axis, and our points have been located as follows—

x	y	Point No
+ 3	+ 2	I
+ 1	- 4	II
- 5	- 6	III
0	+ 2	IV

The co-ordinates of the points are indicated by dotted lines. In practice the graph is plotted upon a grid or network of fine lines, which dispenses with the necessity for inserting dotted lines.

[1] In mathematics the term *curve* includes a *straight line*

A System of Rectangular Co-ordinates

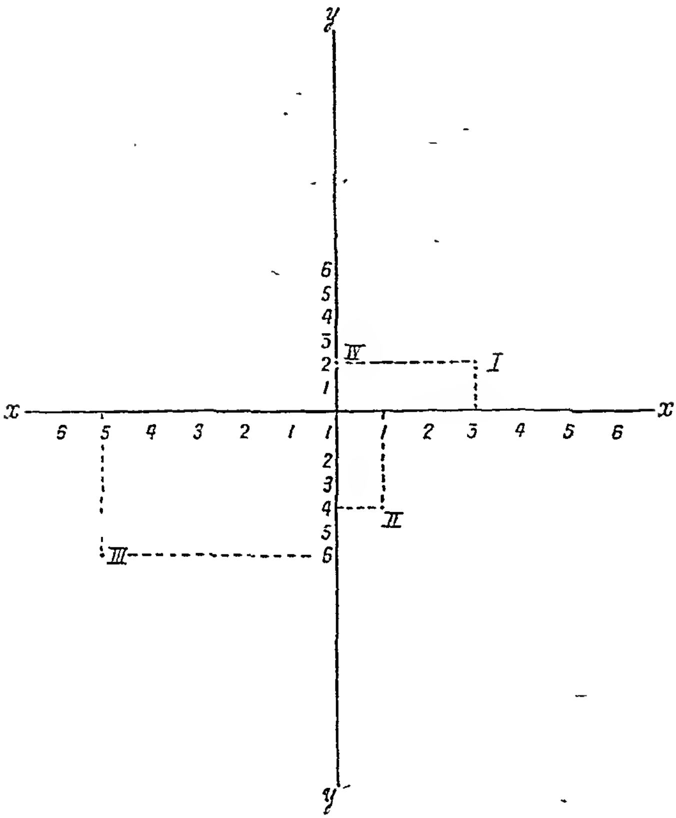


FIG 8

Mathematical Functions.

Graphs may be used to represent mathematical functions, e.g —

$$y = mx + c \qquad \text{(straight line)}$$
$$y = ax^2 + bx + c \qquad \text{(parabola)}$$
$$y = ax^3 + bx^2 + cx + d \qquad \text{(cubic parabola)}$$
$$\left.\begin{array}{r} y = ae^{bx} \\ \text{or } \log y = \log a + bx \end{array}\right\} \text{(compound interest curve)}$$
$$\left.\begin{array}{r} y = ax^b \\ \text{or } \log y = \log a + b \log x \end{array}\right\} \text{(potential curve)}$$

Specimen curves are given in Fig. 9 on p. 38.

As a rule the movements of statistical quantities are too erratic

to admit of *exact* representation by mathematical functions, although *approximate* representation may sometimes be possible.

Specimen Curves

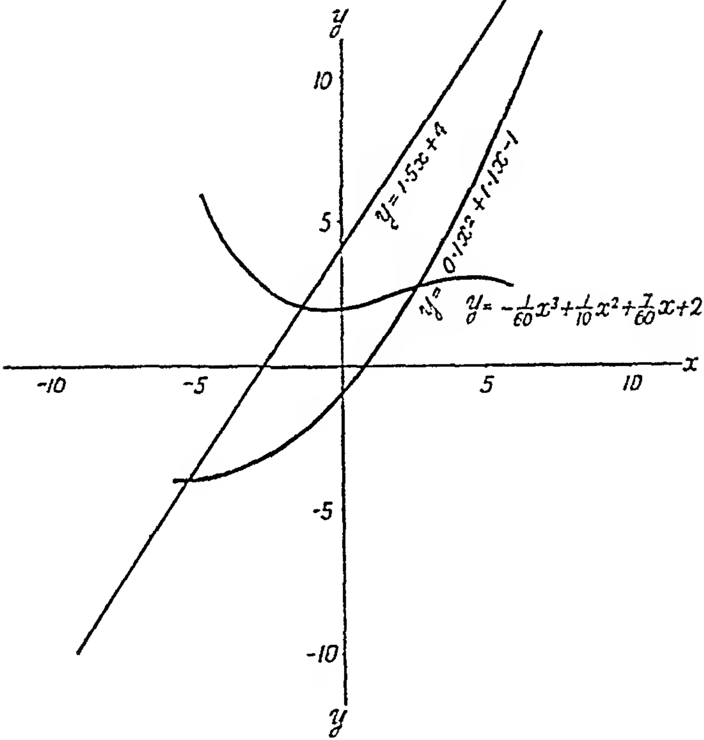


FIG 9

The methods involved are, however, too complicated for an elementary work and they cannot be further discussed here.

Historigrams.

The remainder of this Chapter relates to graphs of temporal series, otherwise called *Historigrams.*[1] Graphs of Frequency Distributions possess special features and are dealt with separately in Chapter IX.

[1] Distinguish carefully between *Historigram* and *Histogram* (see Chapter IX, p 64)

Fig. 10 exhibits the world production of steel over the period 1900–29. One division represents on the x axis five years, and on the y axis 20,000,000 tons of steel. These scales give an outline which is neither too angular nor too flat. Since the quantities

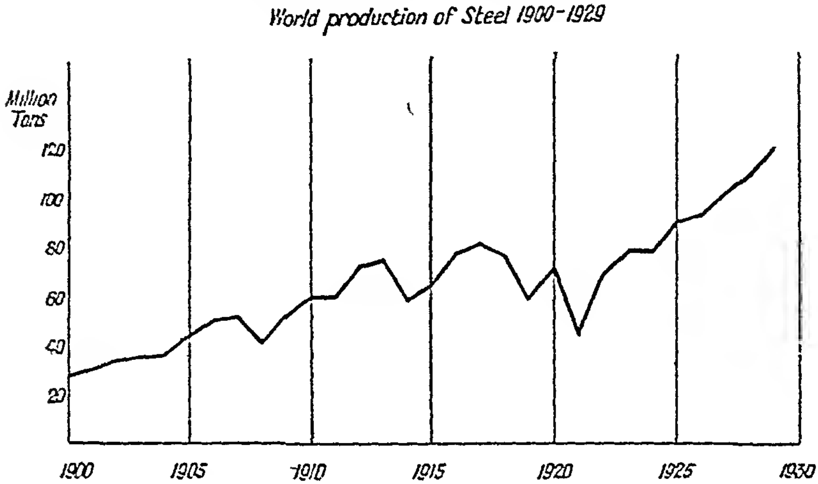


FIG. 10

represented are all positive, only the north-east quadrant of the system of rectangular co-ordinates is necessary.

It is desirable, where space permits, to give the actual figures as well.

Alternatively, the data could have been presented in the form of a bar chart. (See Fig. 11.) The bar chart form is preferred when the data are few in number and irregularly spaced (*Cf.* Fig. 1).

Two or More Variates.

Two or more Variates may be exhibited on the same Graph. Coloured lines may be used where there is any risk of confusion, or if these are precluded on the score of expense, the lines may be differently characterized.

A simple example involving two variates is given in the upper section of Fig. 12.

" Surface " and " Strata " Charts

When the space below the curve is filled in with some kind of

Method I. We convert both series into indices with base = 1924. We take 1924 as base because it is the first of the series and there is no special reason for choosing any other.

Method II. We convert into indices with base = the average of the period 1924–35. This is equivalent to dividing through by 9·77 and 311 respectively and shifting the decimal point two places to the right.

Method III. We choose the scales so that the six largest (smallest) values of A lie on average at the same level as the six largest (smallest) values of B. Assume the scales of A and B are connected by a linear equation—$B = mA + c$. Using the values shown at the foot of the table, we have

$$10{\cdot}10m + c = 325$$

$$9{\cdot}44m + c = 298$$

Whence $m = 40{\cdot}91$ and $c = -88{\cdot}19$.

These precise values are awkward and we accordingly round them off to $m = 40$ and $c = -88$, which gives a neat arrangement. The assumption involved is that a movement of 1 million in employment corresponds with a movement of 40 thousand in marriages. We have in fact plotted the variates so as to equate their mean deviations.[1]

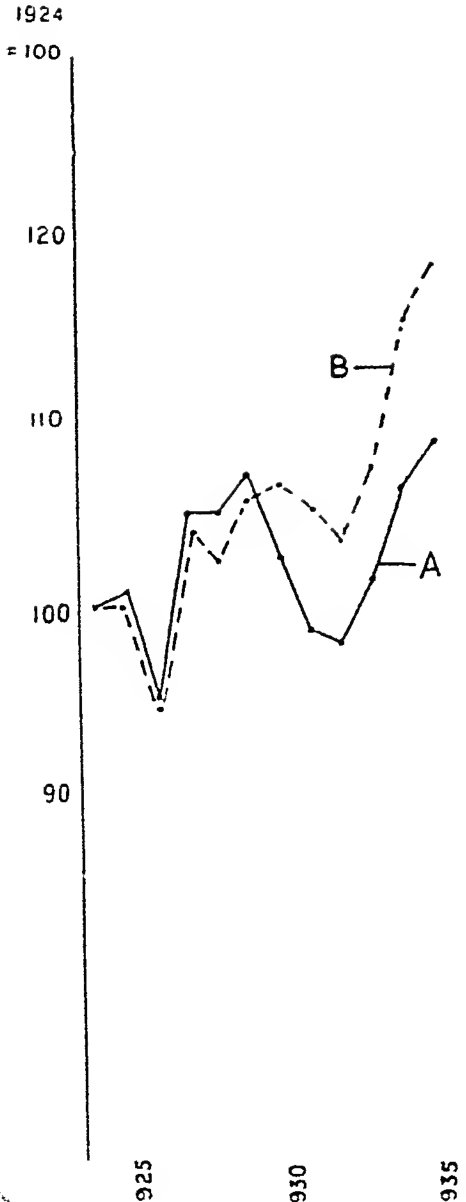
Method IV. We plot both series on a ratio scale,[2] lifting the lower curve up until it lies close to the upper. This operation can be performed either (*a*) by tabulating the logarithms of the two series, adding a constant quantity (say 0·5) to B or (*b*) by plotting them on semi-logarithmic paper as they stand and then raising B a fixed vertical distance by means of a pair of dividers or graduated ruler.

All four methods are arbitrary and there is not much to choose between them in this case. No. I is the easiest and most straightforward since the curves can be continued indefinitely without recalculation of scales. Nos. II, III, and IV give somewhat better comparisons but are more troublesome to calculate. (See Fig. 14.)

Ratio Scales.

So far we have been dealing with natural scale graphs, i.e. graphs

[1] See Chapter XI, p. 98 [2] See next Section.



(B.3133)

TABLE 6

PLOTTING IN UNITS OF DIFFERENT KINDS

Year			Method I		Method II		Method IV	
(1)	A (2)	B (3)	A (4)	B (5)	A (6)	B (7)	A (8)	B (9)
	Mill.	Thous.						
1924	9·53	296	100·0	100·0	97·5	95·2	·979	·971
1925	9·61	296	100·8	100·0	98·4	95·2	·983	·971
1926	9·06	280	95·1	94·6	92·7	90·0	·957	·947
1927	10·02	308	105·1	104·1	102·6	99·0	1·001	·989
1928	10·02	303	105·1	102·4	102·6	97·4	1·001	·981
1929	10·22	313	107·2	105·8	104·6	100·6	1·009	·996
1930	9·80	315	102·8	106·5	100·3	101·3	·991	·998
1931	9·42	312	98·8	105·4	96·4	100·3	·974	·994
1932	9·35	307	98·1	103·7	95·7	98·7	·971	·987
1933	9·68	318	101·6	107·5	99·1	102·3	·986	1·002
1934	10·14	342	106·4	115·6	103·8	110·0	1·006	1·034
1935	10·38	350	108·9	118·3	106·2	112·5	1·016	1·044
Average—6 highest	10·10	325						
Average—6 lowest	9·44	298						
Average overall	9·77	311						

in which the y's are scaled proportionately to their actual values. This method brings out absolute movements in a statistical series but fails to exhibit relative movements in their true light. Thus an increase of £1,000 in the profits of a business is represented by the same vertical distance, whether the average level of profits happen, to be £2,000 or £2,000,000 a year.

The Ratio Scale is employed as an alternative to the Natural Scale whenever it is desired to study relative movements.

With the Natural Scale equal vertical distances represent equal absolute movements; with the Ratio Scale they represent equal proportionate movements.[1]

An absolute series may be converted into a ratio series by plotting either (1) the logarithms of the y's instead of the y's themselves, or (2) the y's themselves upon semi-logarithmic paper (paper ruled

[1] Assume an *increase* of $33\frac{1}{3}$ per cent is represented by 1 in. measured in an *upward* direceion. This will be true all over the graph. It follows that a *decrease* of 25 per cent will be measured by one inch in a *downward* direction.

with a special grid in which vertical distances are scaled logarithmically).

Such a graph is called a semi-logarithmic graph, because one variable appears on a logarithmic whilst the other remains on the natural scale.

Comparison of Natural Scale and Ratio Scale Graphs.

Natural Scale	Ratio Scale
1. Equal vertical distances represent equal *absolute* changes Lines of equal slope represent equal rates of change A straight line denotes a continuous increase at simple interest.	Equal vertical distances represent equal *proportional* changes Lines of equal slope represent equal *proportional* rates of change A straight line denotes a continuous increase at compound interest
2. Suitable for analysing an aggregate into its constituents	Not suitable.
3 Zero and negative values may be shown	These cannot be shown
4 Must be definitely located with reference to a base line, whether shown on the graph or not.	No base line. The whole curve may be moved up and down without affecting its properties.
5 Not suitable when the y variable shows a great range of variation	Eminently suitable when such is the case.

Figure 15 has been drawn so as to illustrate the above propositions

The line AA' represents a compound interest curve rising at the rate of 15 per cent per annum. The ratio scale graph shows this as a straight line.

The line BB' repeats a characteristic feature three times (viz rises of 40, 100, and 50 units followed by a fall of 100 units).

A rise (or fall) of given absolute amount becomes relatively less important as the curve mounts, and this fact is represented in the ratio-scale graph by a progressive flattening of the curve. The necessary numerical details are given in Table 7, on p 45, in order that the student may be in a position to check them.

The paper used in Fig. 15 is known as "three cycle" paper. This provides for a range of values represented by maximum = 1000 × minimum Rulings are available up to six cycles.

Fig 16 shown on p. 47 supplies an instance of a series that could not possibly be plotted except upon a ratio scale.

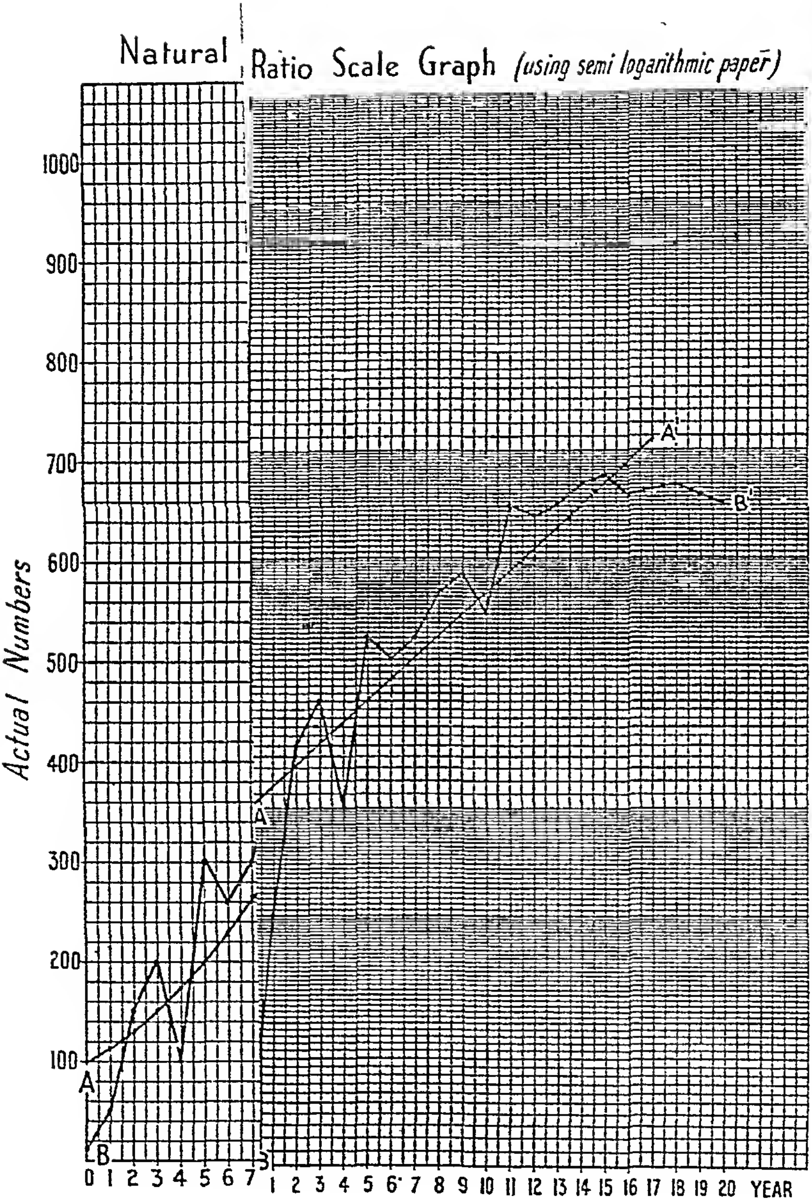

Natural
Ratio Scale Graph (using semi logarithmic paper)
Actual Numbers
1000
900
800
700
600
500
400
300
200
100
A
B
A
B
A'
B'
0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 YEAR

TABLE 7

DETAILS SUPPORTING FIG. 15

Year (1)	Line AA' (y) (2)	Log y (3)	Line BB' (z) (4)	Log z (5)
0	100	2·0000	10	1·0000
1	115	2·0607	50	1·6990
2	132·2	2·1214	150	2·1761
3	152·1	2·1821	200	2·3010
4	175·3	2·2428	100	2·0000
5	201·1	2·3035	300	2·4771
6	231·3	2·3642	260	2·4150
7	266·1	2·4249	300	2·4771
8	305·9	2·4856	400	2·6020
9	351·8	2·5463	450	2·6532
10	404·6	2·6070	350	2·5441
11	465·2	2·6677	700	2·8451
12	535·1	2·7284	660	2·8195
13	615·3	2·7891	700	2·8451
14	707·6	2·8498	800	2·9031
15	813·7	2·9105	850	2·9294
16	035·8	2·9712	750	2·8751
17	1,076	3·0319	775	2·8893
18	—	—	800	2·9031
19	—	—	750	2·8751
20	—	—	700	2·8451

Analysis of Time Series.

Developments upon the economic side have brought the question of the analysis of Time Series into prominence. In view of the complexity of the problem and its controversial nature, only a brief summary can here be attempted.

It is generally supposed that an economic series involves five elements, viz.—

1. The Trend or course that would be taken by the curve in the absence of disturbing factors.

2. Cyclical fluctuations[1] or wave-like disturbances corresponding with the movements of the Trade Cycle, which is usually supposed to extend over a period of 7 to 10 years.

3. Seasonal variations associated with the harvests, the weather,

[1] This analysis applies to long periods (say, 30 years). Over a short period (say, 10 years) it is usually impossible to separate the trend from the cycles In this case we may fit a composite curve known as the Trend Cycle.

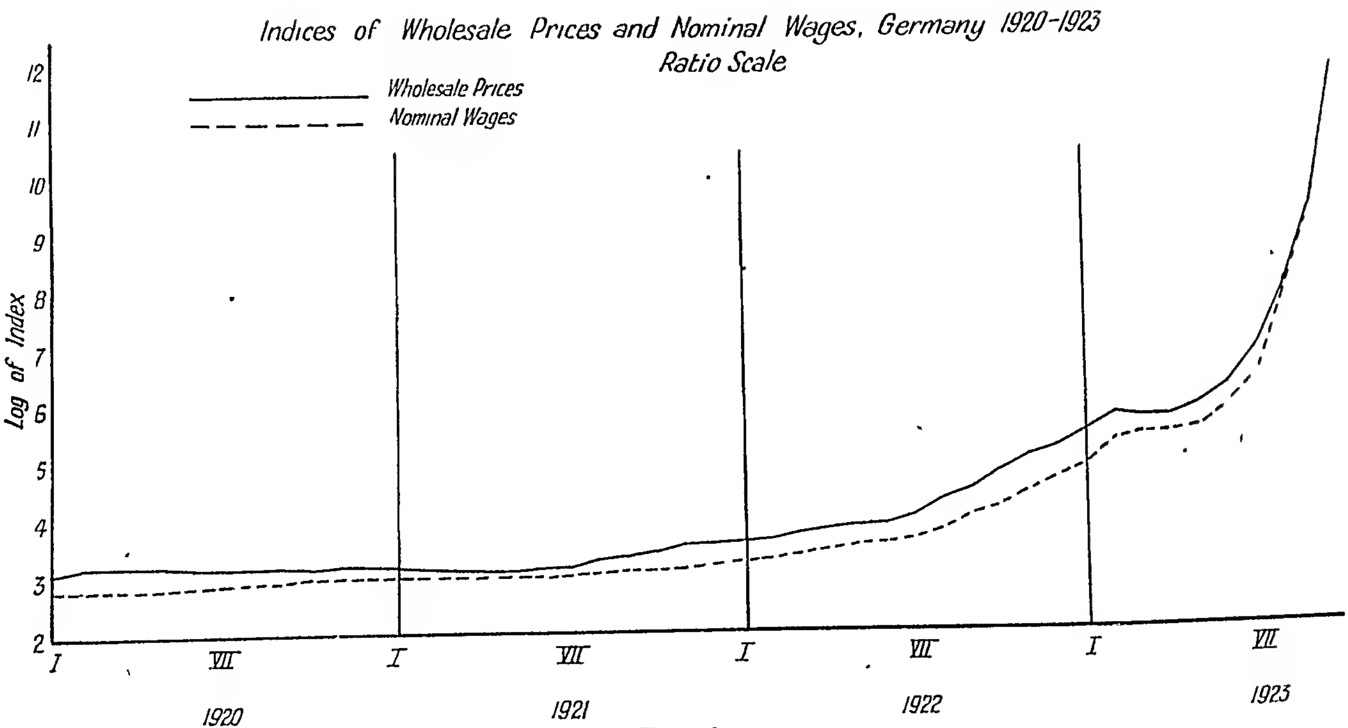

Indices of Wholesale Prices and Nominal Wages, Germany 1920-1923
Ratio Scale
Wholesale Prices
Nominal Wages
Log of Index
12
11
10
9
8
7
6
5
4
3
2
I
VII
I
VII
I
VII
I
VII
1920
1921
1922
1923

FIG. 16

the varying length of the month and other annually recurrent phenomena.

4. Catastrophic movements caused by unusual or unexpected events.

5. Residuals, which include all movements not included under heads (1) to (4).

Various methods, not always very successful, have been devised with a view to breaking up series into the five elements mentioned. In this section we will show how the trend can be separated from the rest of the series in a simple case.

The first step is to plot the series, upon either a natural or a ratio scale. Assuming then that the fluctuations from the trend are in the long run as likely to be positive as negative, the trend may be approximately located by any of the following methods—

1. Drawing a smooth curve freehand through the observations upon a plan that ignores minor disturbances but gives expression to major disturbances of known origin.

2. Employing the method of the moving average.

3. Fitting a mathematical curve to the observations.[1]

Method (1) demands a higher degree of skill and judgment than is possessed by the average beginner, and even trained operators may differ considerably in the manner in which they locate their curves.

TABLE 9

ANNUAL PERCENTAGES UNEMPLOYED AMONG MEMBERS OF CERTAIN TRADE UNIONS, 1881–1920[1]

Year (1)	Yearly Average of Percentages Unemployed at End of each Month (2)	Nine Year Moving Average (3)	Col. (3) Adjusted (4)	Col. (2)–Col. (4) (5)
1881	3·5	—	6·6	− 3·1
1882	2·3	—	6·6	− 4·3
1883	2·6	—	6·5	− 3·9
1884	8·1	—	6·4	+ 1·7
1885	9·3	5·6	6·2	+ 3·1
1886	10·2	5·5	6·0	+ 4·2
1887	7·6	5·6	5·7	+ 1·9
1888	4·9	6·0	5·4	− 0·5
1889	2·1	5·9	5·2	− 3·1
1890	2·1	5·7	5·0	− 2·9
1891	3·5	5·2	4·8	− 1·3
1892	6·3	4·7	4·6	+ 1·7
1893	7·5	4·5	4·4	+ 3·1
1894	6·9	4·6	4·2	+ 2·7
1895	5·8	4·6	4·0	+ 1·8
1896	3·3	4·5	3·9	− 0·6
1897	3·3	4·2	3·8	− 0·5
1898	2·8	3·8	3·7	− 0·9
1899	2·0	3·5	3·6	− 1·6
1900	2·5	3·5	3·6	− 1·1
1901	3·3	3·7	3·7	− 0·4
1902	4·0	3·8	3·8	+ 0·2
1903	4·7	3·9	4·1	+ 0·6
1904	6·0	4·5	4·5	+ 1·5
1905	5·0	5·1	4·8	+ 0·2
1906	3·6	5·2	5·0	− 1·4
1907	3·7	5·1	4·9	− 1·2
1908	7·8	5·0	4·8	+ 3·0
1909	7·7	4·5	4·6	+ 3·1
1910	4·7	4·3	4·3	+ 0·4
1911	3·0	4·1	4·0	− 1·0
1912	3·2	3·7	3·6	− 0·4
1913	2·1	2·9	3·3	− 1·2
1914	3·3	2·1	2·6	+ 0·7
1915	1·1	1·9	2·1	− 1·0
1916	0·4	1·8	1·6	− 1·2
1917	0·7	—	1·3	− 0·6
1918	0·8	—	1·3	− 0·5
1919	2·4	—	1·6	+ 0·8
1920	2·4	—	3·0	− 0·6

[1] *Twentieth Abstract of Labour Statistics of the United Kingdom*, Cmd. 3831 (1931), p. 72.

relates, instead of against the central year. This plan saves the trouble of centring, but is unsuitable when the deviations have to be calculated.

The moving average gives a satisfactory representation of trend when the data lie approximately on a straight line. Owing, however, to its property of cutting corners it is not so suitable when they lie on a curve. Methods of overcoming this difficulty are available, but involve highly complicated calculations.

Seasonal Variations.

When a series is strongly seasonal (e.g. employment in the building industry, retail trade), the fluctuations are apt to be confusing, and various methods have been devised for removing them. The following is the most straightforward—

1. We take a twelve-month moving average of the data. This incorporates the effect of (*a*) the long period trend and (*b*) the cycle. We call this the Trend-cycle to distinguish it from the Trend proper, which is understood to apply to a considerable period of years.

2. We then find the monthly deviations of the original series from the trend-cycle and average them for each month separately. The effect of averaging by months is to bring out items common to the respective months, i.e. the seasonal variations, and to merge the remainder.

3. These monthly averages are adjusted so as to add up to zero. In this form they provide the adjustments required.

In the following example the method has been applied to quarterly data of industrial production. The use of quarterly data does not affect the principle, but simplifies the arithmetic.

Column (2) shows the quarterly figures of industrial production as published by the *Board of Trade* (after adjustment throughout to the year 1930 as base). To find the centred moving average of four items we sum column (2) in fours, placing the result in column (3) against the second item, and then sum in pairs, placing the result in column (4) against the third item. Finally we divide by 8, placing the quotient in column (5). This gives the moving average properly centred.[1]

TABLE 10

Industrial Production—United Kingdom
Board of Trade Index—All Items
(1930 = 100)

Year and Quarter	Index of Production (1930 = 100)	Sum in Fours	Sum in Pairs	Divide by 8	Difference Col. (2)–Col. (5)	Normal Seasonal Movement from Table 11	Index of Production Adjusted for Seasonal Change Col. (2)–Col. (7)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1928							
1 .	106·0	—	—	—	—	+ 1·8	104·2
2 .	100·4	408·6	—	—	—	− 0·5	100·9
3 .	97·1	409·8	818·4	102·3	− 5·2	− 3·9	101·0
4 .	105·1	418·0	827·8	103·5	+ 1·6	+ 2·6	102·5
1929							
1 .	107·2	428·2	846·2	105·8	+ 1·4	+ 1·8	105·4
2 .	108·6	433·6	861·8	107·7	+ 0·9	− 0·5	109·1
3 .	107·3	434·0	867·6	108·4	− 1·1	− 3·9	111·2
4 .	110·5	425·4	859·4	107·4	+ 3·1	+ 2·6	107·9
1930							
1 .	107·6	414·6	840·0	105·0	+ 2·6	+ 1·8	105·8
2 .	100·0	400·1	814·7	101·8	− 1·8	− 0·5	100·5
3 .	96·5	384·0	784·1	98·0	− 1·5	− 3·9	100·4
4 .	96·0	373·1	757·1	94·6	+ 1·4	+ 2·6	93·4
1931							
1 .	91·5	363·0	736·1	92·0	− 0·5	+ 1·8	89·7
2 .	89·1	361·1	724·1	90·5	− 1·4	− 0·5	89·6
3 .	86·4	361·3	722·4	90·3	− 3·9	− 3·9	90·3
4 .	94·1	363·2	724·5	90·6	+ 3·5	+ 2·6	91·5
1932							
1 .	91·7	361·2	724·4	90·6	+ 1·1	+ 1·8	89·9
2 .	91·0	358·8	720·0	90·0	+ 1·0	− 0·5	91·5
3 .	84·4	358·4	717·2	89·6	− 5·2	− 3·9	88·3
4 .	91·7	360·5	718·9	89·9	+ 1·8	+ 2·6	89·1
1933							
1 .	91·3	369·3	729·8	91·2	+ 0·1	+ 1·8	89·5
2 .	93·1	378·7	748·0	93·5	− 0·4	− 0·5	93·6
3 .	93·2	393·1	771·8	96·5	− 3·3	− 3·9	97·1
4 .	101·1	404·6	797·7	99·7	+ 1·4	+ 2·6	98·5
1934							
1 .	105·7	414·6	819·2	102·4	+ 3·3	+ 1·8	103·9
2 .	104·6	425·4	840·0	105·0	− 0·4	− 0·5	105·1
3 .	103·2	432·7	858·1	107·3	− 4·1	− 3·9	107·1
4 .	111·9	439·6	872·3	109·0	+ 2·9	+ 2·6	109·3
1935							
1 .	113·0	447·1	886·7	110·8	+ 2·2	+ 1·8	111·2
2 .	111·5	455·9	903·0	112·9	− 1·4	− 0·5	112·0
3 .	110·7	466·1	922·0	115·2	− 4·5	− 3·9	114·6
4 .	120·7	478·0	944·1	118·0	+ 2·7	+ 2·6	118·1
1936							
1 .	123·2	489·7	967·7	121·0	+ 2·2	+ 1·8	121·4
2 .	123·4	501·1	990·8	123·8	− 0·4	− 0·5	123·9
3 .	122·4	509·8	1010·9	126·4	− 4·0	− 3·9	126·3
4 .	132·1	521·1	1030·9	128·9	+ 3·2	+ 2·6	129·5
1937							
1 .	131·9	—	—	—	—	+ 1·8	130·1
2 .	134·7	—	—	—	—	− 0·5	135·2

Column (6) shows the quarterly deviations from the moving average. These are taken out and tabulated in the form shown below. The values in the table fluctuate considerably, and in order to eliminate the extremes, it is proposed to take the extended median[1] (found by averaging the four middle items for the first two quarters and the three middle ones for the remaining two). The sum of the *four medians is — 0·4 and this is adjusted to zero by adding 0·1* unit to each item. The normal seasonal movements so determined are entered in column (7) of Table 10. Deducting them from column (2) gives column (8), which represents the original data adjusted for seasonal changes. It is left as an exercise for the student to plot columns (2) and (8). He will observe that the adjusted series runs more smoothly than the original, and that little sign of the seasons is *left*.

Having mastered the *principle*, the student should *try* it out on some monthly series with strong seasonal movements, e.g. unemployment or foreign trade.

Possible improvements[2] in the method include, (1) expressing the

seasonal movements as percentages of the trend-cycle instead of as differences, (2) using variable adjustments instead of fixed ones, and (3) calculating the seasonal variations collectively by the "short-cut" method. Further elaborations, including the well-known "Link relative" method, are too complicated to be discussed here.

The reader will notice that this method allows automatically for the varying length of the month as well as for public holidays with exception of Easter. It does not allow for the varying number of Sundays in the month, nor for the incidence of Easter, nor for movable factors associated with the seasons but not strictly tied down to particular months, e.g. purchases of summer clothing.

Rules for Construction of Historigrams.

The following are the generally accepted rules—

1. The general arrangement should proceed from left to right.

2. Where possible, represent quantities by linear magnitudes, as areas or volumes are more likely to be misinterpreted.

3. For a curve, the vertical scale, whenever practicable, should be so selected that the zero line will appear on the chart.

4. If the zero line of the vertical scale will not normally appear on the curve diagram, the zero line should be shown by the use of a horizontal break in the diagram.[1]

5. The zero lines of the scales for a curve should be sharply distinguished from the other co-ordinate lines.

6. For curves having a scale representing percentages, it is usually desirable to emphasize in some distinctive way the 100 per cent line or other line used as a basis of comparison.

7. When the scale of a chart refers to dates, and the period represented is not a complete unit, it is better not to emphasize the first and last ordinates, since such a chart does not represent the beginning nor end of time.

8. It is advisable not to show any more co-ordinate lines than necessary to guide the eye in reading the chart.

9. The curve lines of a chart should be sharply distinguished from the ruling.

[1] Rules (4) is troublesome to observe when we are using commercial graph paper; consequently it is often ignored. Provided the scale is boldly marked along the margin of the graph, little is lost by ignoring it.

10. In curves representing a series of observations, it is advisable whenever possible, to indicate clearly on the chart all the points representing the separate observations.

11. The horizontal scale for curves should usually read from left to right and the vertical scale from bottom to top.

12. Figures for the scales of a chart should be placed at the left and at the bottom, or along the respective axes.

13. It is often desirable to include in the chart the numerical data or formulas represented.

14. If numerical data are not included in the chart, it is desirable to give the data in tabular form accompanying the chart.

15. All lettering and all figures should be placed so as to be easily read from the base as the bottom, or from the right-hand edge of the chart as the bottom.

16. The title should be made as clear and complete as possible. Sub-titles or descriptions should be added, if necessary, to ensure clearness.

Code of Preferred Practice for Graphic Presentation.

The American Society of Engineers (Committee on Standards for Graphic Presentation) have recently issued a *Code of Preferred Practice for Graphic Presentation (Time Series Charts)*. The code is being issued in preliminary form for whatever immediate value it may have for interested users and for the purpose of securing their criticisms and suggestions before presenting it for approval and transmission to the *American Standards Association*. Copies of the Code may be obtained from the Society's offices at 29 West 39th Street, New York.

CHAPTER VIII

DERIVATIVE DATA

A Statistical Derivative is a Quantity formed by Combination of Two or More Original Items. Complex functions such as averages and measures of dispersion will be considered in their proper place. This section will be confined to simple derivatives.

The relationships to be expressed are of two kinds—

1. Co-ordinate, involving the combination of quantities of equal standing.

2. Subordinate, involving the comparison of a quantity with the whole of which it forms part.

Co-ordinate Derivatives.

Co-ordinate derivatives include—

1. The Simple Difference between two quantities of the same kind. Thus we may compare this year's expenditure with last year's, or this year's actual with this year's budgeted expenditure.

2. The Percentage Difference. In this case the difference is expressed not as an absolute quantity but as a percentage upon some quantity taken as standard.

3. The Ratio, which is really another way of expressing the percentage difference, e.g.

$$\frac{x_1}{x_0} = 1 + \frac{x_1 - x_0}{x_0} = 1 - \frac{x_0 - x_1}{x_0}$$

A ratio may be written in various fashions, e.g. $15 : 5 = 75 : 25 = 3 : 1 = 1 : \frac{1}{3}$, are identical ways of expressing the same relationship.

4. The Rate, which differs from the ratio in that the numerator and denominator involve quantities of different kinds. We speak of the *ratio* of male births to female births because both quantities are of the same kind (viz. births), but we speak of sickness rates and accident rates because the latter involve comparison of quantities of different kinds (cases with persons).

A rate is usually standardized with respect to the denominator,

e.g. 5 per cent = 50 per mille = 0·05 (per unit). In the last case it is known as a co-efficient.

The distinction between ratios and rates is not perfectly definite. We usually speak of the *death rate* (number of deaths per 1,000 population), but it would be equally correct to speak of the *death ratio* (number of persons dying to number living).

Subordinate Derivatives.

Subordinate derivatives involving the relationship of parts to the whole are expressed in the form of proportions or percentages. We speak of the proportion or percentage of males and females in the total population because it is a question of dividing the total population between those categories.

Statistical derivatives which comply with the conditions laid down in Chapter V, page 17, form a derivative series, which may be tabulated or plotted in the usual way. Derivative series usually aim at the elimination of some disturbing factor which prevents effective comparison. A series representing foreign trade per head of population eliminates the chief effects of demographic changes, and a series representing real wages (i.e. money wages divided by cost of living) eliminates variations in the purchasing power of money. In general the test of a derivative series is its stability, which may be measured graphically or by computing its dispersion.[1] The greater the degree of stability for that class of data, the more reliable are the indications of that series. The subject of stability has received much attention from statisticians, but the problems involved are complicated, and only their simpler issues can be discussed here.

Rules for Forming Derivatives.

1. The first rule for forming a derivative is to secure homogeneity. The so-called crude death rate

$$= \frac{1000 \times \text{number of deaths}}{\text{population}}$$

is useful in its way, but it fails as a satisfactory measure of mortality because the figures are heterogeneous Mortality varies considerably with age and sex, and the employment of a flat death

[1] See Chapter XI, p. 98

rate ignores all differences due to the varying age and sex composition of the population. It is better, therefore, to break up the figures into age groups and calculate the death rate for each separately.

For the same reason it is dangerous to estimate the profits of a business upon the basis of an average rate of profit upon turnover. Different articles carry different rates of profit, and the calculation should reflect the changes in the relative numbers of articles sold.

2. The second rule may be expressed by saying that the quantities compared must cover exactly the same ground. In other words, it is necessary to consider where the risk attaches. A marriage rate should be calculated on the number of single and widowed persons of marriageable age, a legitimate birth rate upon the number of married females of child-bearing age, an accident rate upon the number of persons exposed to risk of accident, and so on.

Difficulties sometimes occur owing to the fact that the number of persons at risk is a varying quantity. A factory pay-roll varies from week to week, and the number of insurable persons in an industry is affected by entrants and exitants. In such case it is usual to calculate an average number exposed to risk. Upon this basis the accident rate at a factory would be represented by the formula

$$1000 \times \frac{\text{No of accidents occurring during the year}}{\frac{1}{52}\text{ (total number of names on 52 weekly pay rolls)}}$$

Classes at Risk.

Doubts frequently arise whether a particular class of person is at risk or not. Fertility of marriage may be calculated in the form

$$\frac{\text{Number of children born}}{\text{Number of marriages of completed fertility}^{1}}$$

Whether sterile marriages should be included or not is a debatable point. A sterile marriage, whether so from physiological causes or from deliberate choice, evidently carries no "risk" of children. On the other hand the laws of sterility are so obscure that differentiation is invidious.

[1] That is, a marriage in which the wife has passed the child-bearing age (conventionally taken at 45 years) The inclusion of marriages of uncompleted fertility would of course vitiate the results

The burden of income tax may be expressed as so much per head. Should, however, the taxation per head be calculated upon—

(*a*) The number of persons in receipt of money incomes; or

(*b*) The number of persons actually paying tax?

Persons who do not pay tax are not exposed to risk of taxation so long as the existing tax laws and regulations remain in force. On the other hand, they may be exposed to future risk in the event of a change in the tax system. It is therefore preferable to calculate upon basis (*a*).

CHAPTER IX

STATISTICAL GROUPS

SINCE their characteristics can be graded as well as described, quantitative variates lend themselves to special methods of treatment that are not available for qualitative variates. Consider the following record of weekly wages taken from an inquiry (pre-war) into working-class conditions—

TABLE 12

STATISTICS OF WAGES OF WEEKLY WAGE-EARNERS
CRUDE DATA

Weekly Wage (1)	No. of Wage-earners (2)	Weekly Wage (1)	No. of Wage-earners (2)
s d		s d	
14 –	1	28 –	1
15 –	1	29 –	1
18 –	4	30 –	10
19 –	2	31 –	1
20 –	7	32 –	1
20 6	1	32 6	1
21 –	4	35 –	1
22 –	4	36 –	1
23 –	2	38 –	1
24 –	8	40 –	3
25 –	7	45 –	6
25 6	1	50 –	1
27 –	1	55 –	1
		TOTAL .	72

The record includes seventy-two wage-earners with weekly wages varying from 14s. to 55s The quantitative elements involved are two in number, viz.—

1. Measurable Characteristic. Weekly wage.
2. Frequency. Number of earners with given wage.

The figures as they stand involve too much detail for ready comprehension, and the information must therefore be condensed and summarized.

The principle of classification implies that the wage-earners

should be arranged in groups according to their earning capacity, and there are two methods of achieving this object—

1. **The Frequency Distribution.** Wages graded in equal intervals involving groups of unequal size.

2. **The Grouped Array.** Groups of equal size, involving wage intervals of unequal size.

Frequency Distribution.

In the following table the grouping starts at 12s. 6d. and proceeds by equal intervals of 5s. to a maximum of 57s. 6d., giving eight groups in all. A broader grading (say by 10s.) would have been uninformative, and a narrower grading (say by 4s.) would have caused irregularity in the resultant figures. Since most of the observations fall on multiples of one shilling, the divisions between grades are arranged to fall on multiples of sixpence. This arrangement secures an even distribution of the observations within the grades.

Alternative Methods.

The following are alternative methods of describing the group intervals—

s.	s.	
12·5 —	17·49	
17·5 —	22·49	(A)
12·5 —	17·5	
17·5 —	22·5	(B)
12·5 —		
17·5 —	. . . . .	(C)

Whichever method is adopted, it must be made quite clear what is to be done with marginal cases. Method (B) in particular is faulty in this respect. Frequently doubtful cases can be decided by taking the calculations to a further place of decimals. If other devices fail, marginal items may be divided, one-half being assigned to the group above and the other half to the group below.

The Central Wage.

For the Purposes of this Analysis the Wage-earners in (say) the 17s. 6d. to 22s. 6d. group, are treated as if they all earned the same wage. This may be taken (with slight loss of accuracy) as 20s., and the table can be rewritten as follows—

TABLE 14

FREQUENCY DISTRIBUTION OF WAGES OF WEEKLY WAGE-EARNERS—II

Central Wage (1)	No of Wage-earners (2)
s.	
15	2
20	22
25	19
30	14
35	3
40	4
45	6
50	1
55	1
TOTAL .	72

Alternative Grouping.

Let us now test the effect of a different system of grouping. In the following table the groups begin at 12s. and proceed by intervals of 8s —

TABLE 15

FREQUENCY DISTRIBUTION OF WAGES OF WEEKLY WAGE-EARNERS—III

Weekly Wage (1)		Central Wage (2)	No of Wage-earners (3)
	and under		
s	s	s	
12	20	16	8
20	28	24	35
28	36	32	16
36	44	40	5
44	52	48	7
52	60	56	1
		TOTAL	72

Table 13 can be illustrated graphically either as a Histogram (Fig. 19) or as a Frequency Polygon (Fig 20).

Fig. 19. Histogram.

The group intervals are plotted along the x axis, and on each division is drawn a rectangle with *area* proportional to the number of observations recorded in the table. The chart has been scaled so that one square represents 5 wage earners The second rectangle covers 4·4 squares and $4{\cdot}4 \times 5 = 22$.

This form of graph is known as a Histogram, block diagram, or stair-case chart The dotted line should be ignored at this stage

Fig. 20. Frequency Polygon.

Alternatively, the graph may be completed as a Frequency Polygon. In this form of construction the notion of area is pushed into the background. Points are drawn over the centres of class intervals at distances proportional to the frequencies, and are connected by straight lines The frequency polygon has been scaled so that one large division represents 5 wage earners The third point from the left is distant 4·4 divisions from the x axis and $4{\cdot}4 \times 5 = 22$. Note

Frequency Distribution of Wages of Weekly Wage-earners, Histogram illustrating Table 13
The dotted line shows the smoothed curve.

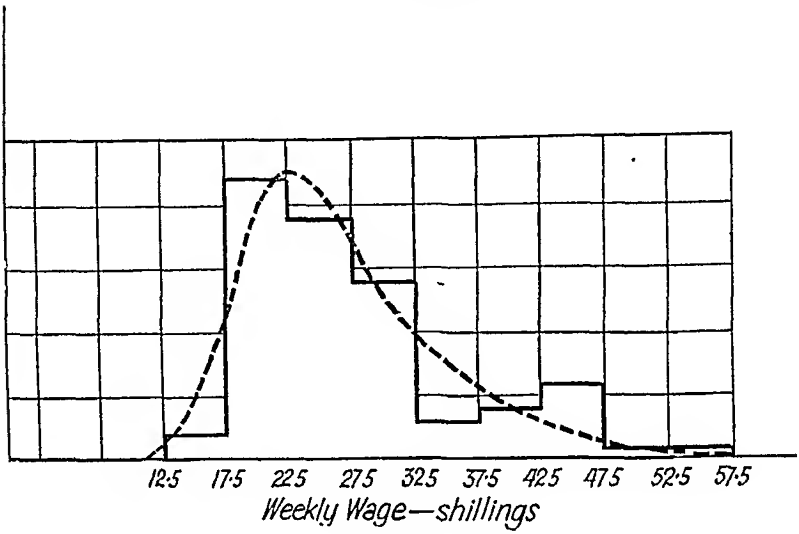


FIG. 19

Frequency Distribution of Wages of Weekly Wage-earners Frequency Polygon illustrating Table 13

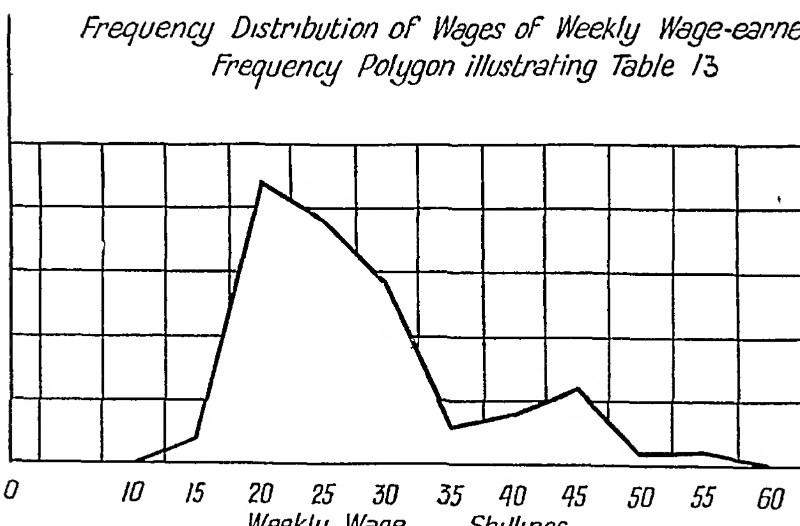


FIG. 20

that the terminals are located at 10s and 60s (not 12·5s and 57·5s) so that the total area of the polygon is equivalent to that of the corresponding Histogram

Histograms Illustrating Alternative Grouping

The following figure illustrates the system of grouping by intervals of 8s shown in Table 15. For purposes of comparison Fig 19 is

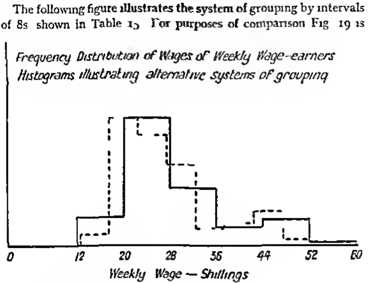


Fig 21

shown by means of dotted lines. The effect of the broader grouping is clearly brought out

Advantages of the Histogram Construction

1 By its dependence upon the notion of area it gives expression to the distribution of the total amongst a number of groups

2 It is easier to smooth than the frequency polygon

3 It can be used with unequal group intervals

Advantages of the Frequency Polygon

1 It is suitable for comparative purposes. Two or more polygons may be plotted on the same chart because the lines tend to cross and not to overlap

2 It is more readily understood by the layman

Histogram with Unequal Group Intervals.

Ideally, all the Group Intervals should be equal in width. In practice, however, data are frequently tabulated by unequal intervals, either because more detail is required on one part of the scale than on another, or else in order to save cost of printing. The following table affords an example—

TABLE 16

GREAT BRITAIN—DISTRIBUTION OF GAINFULLY OCCUPIED MALE POPULATION BY AGES, 1921[1]

Age (1)		No. of Males Gainfully Occupied (2)	No. per Ten Year Interval (3)
Years	and not exceeding Years	(000's)	(000's)
12	14	44	220
14	16	532	2,660
16	18	725	3,625
18	20	739	3,695
20	25	1,601	3,202
25	35	2,887	2,887
35	45	2,731	2,731
45	55	2,318	2,318
55	65	1,429	1,429
65	70	404	808
70	—	246	492 (?)
	TOTAL	13,656	

$a = 37{\cdot}61$ years; $M = 36{\cdot}10$ years; $Z = 18{\cdot}60$ years.
$D_1 = 18{\cdot}17$ years; $Q_1 = 24{\cdot}29$ years; $Q_3 = 49{\cdot}24$ years; $D_9 = 59{\cdot}99$ years.
$\eta = 12{\cdot}88$ years; $\sigma = 15{\cdot}17$ years; $QD = 12{\cdot}47$ years.
$CV = 40{\cdot}35$ per cent; $j_1 = 1{\cdot}23$.

There are	4	groups of	2 years
,,	1	,,	5 ,,
,,	4	,,	10 ,,
,,	1	,,	5 ,,
,,	1	,,	indeterminate.

Plotting the graph is assisted by the addition of a third column showing the number of persons per ten-year interval. Thus a density of 532,000 persons per two-year interval (14–16) corresponds

[1] *Twentieth Abstract of Labour Statistics for the United Kingdom*, 1931 (Cmd. 3831), p. 3.

with a density of 2 660 000 persons per ten year interval The heights of the rectangles in the graph are drawn proportionately to column (3) Upon this basis one square represents 500 000 persons [1] No great error is involved by the assumption that the density of the last group is 492

The polygon construction is not suitable in this case

Smoothing

Provided the material is homogeneous [2] and the number of items sufficiently large the histogram will usually show regular tendencies The typical frequency distribution is uni modal i e rising from zero to a high peak and then falling to zero again If the distribution shows two or more modes there is a presumption that the material is heterogeneous Narrowing the class interval will increase the regularity of the histogram up to a point then irregularities will break out

In general the larger the number of items the finer may be the grouping and the smoother will be the appearance of the figure With an indefinitely large number of items and indefinitely small intervals a regular histogram will merge into a smooth curve

Upon this basis there is justification for smoothing the Histogram as it stands Smoothing consists in drawing a regular curve through the figure rounding off angles in such manner that—

(*a*) The total area of the smoothed figure is exactly equal to the total area of the original figure

(*b*) The area subtended by each segment of the smoothed curve is approximately equal to that of the corresponding rectangle Note that the peak of the curve rises above the peak of the histogram

In Fig 19 (relating to wages of wage earners) the Histogram has been smoothed as indicated by the dotted line Since the number of items (72) is small it is not thought that the secondary peak indicated by the seventh group interval is statistically significant and therefore it has been smoothed out

Fig 22 (relating to ages of the gainfully occupied population) has also been smoothed on similar principles There is a depression

[1] The area of the second rectangle is

5 32 × 0 2 divisions — 1 064 squares
— 532 000 persons

[2] Data are said to be *homogeneous* when they are alike in relevant aspects and *heterogeneous* when they are not alike

between years 20–40, corresponding to war losses. This depression is statistically significant, and the curve has accordingly been flexed inwards so as to give expression to this fact.

Great Britain — Distribution of gainfully occupied Male population by ages, 1921.
(The dotted line denotes the smoothed curve)

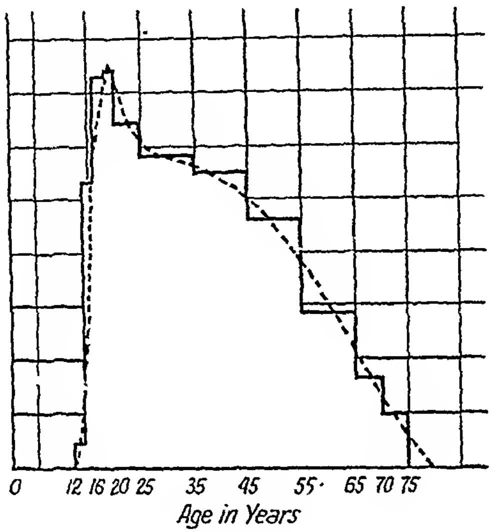


FIG. 22

Theory of Smoothing.

The theory underlying the smoothing process is that the given (finite) distribution forms a random sample[1] of an indefinitely large population obeying a regular law, and that the departures from regularity (fluctuations) shown by the sample distribution are due to its smallness, which tends to exaggerate the influence of abnormal items. The smoothed curve based upon the histogram represents (approximately at least) the ideal distribution that would be exhibited by the totality of similar data, were we in a position to study it. Smoothing gives expression to the underlying regularity and unity of phenomena.

Smoothing is less easy than it seems. There is the technical difficulty of drawing a regular curve of the same area as the

[1] See Chapter XIV.

histogram and there is also the difficulty of deciding whether a given irregularity is statistically significant in which case it must be left in or not significant in which case it must be smoothed out Various mathematical methods are available but they depend for the most part upon advanced theorems lying outside the scope of this work [1]

Finally there are no grounds for smoothing a histogram unless it is fairly regular to begin with and it is supposed that the data exemplify a general law of distribution

Cumulative Frequency Distribution (Ogive)

For some purposes Cumulative Frequency Distributions (Ogives) are more useful than ordinary (or non-cumulative) distributions Let us rewrite Table 13 as follows—

TABLE 17

CUMULATIVE FREQUENCY DISTRIBUTION OF WAGES OF WEEKLY WAGE EARNERS

Weekly Wage ( )		No of Wage earners (2)	Cumulative No (3)
	and under		
s	s		
12 5	17 5	2	2
17 5	22 5	22	24
22 5	27 5	19	43
27 5	32 5	14	57
32 5	37 5	3	60
37 5	42 5	4	64
42 5	47 5	6	70
47 5	52 5	1	71
52 5	57 5	1	72
	TOTAL	72	

Column (3) of this table reads that there were two wage earners with wages under 17 5s twenty four with wages under 22 5s etc etc Fig 23 exhibits the result of plotting column (3) whilst Fig 24 exhibits the result of cumulating in reverse order

The advantages of the Ogive are that it runs more regularly than the non cumulative figure and that little difficulty is caused by variations in the group intervals The ogive may be smoothed as indicated by the dotted line

[1] See however Chapter XVII p 190 for an elementary method

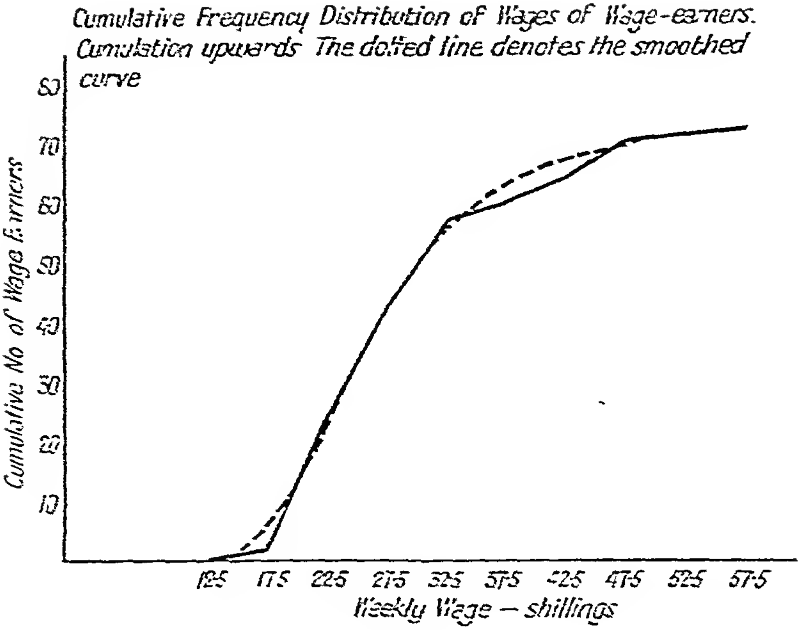

Cumulative Frequency Distribution of Wages of Wage-earners. Cumulation upwards The dotted line denotes the smoothed curve
Cumulative No of Wage Earners
10
20
30
40
50
60
70
80
12·5
17·5
22·5
27·5
32·5
37·5
42·5
47·5
52·5
57·5
Weekly Wage — shillings

FIG. 23

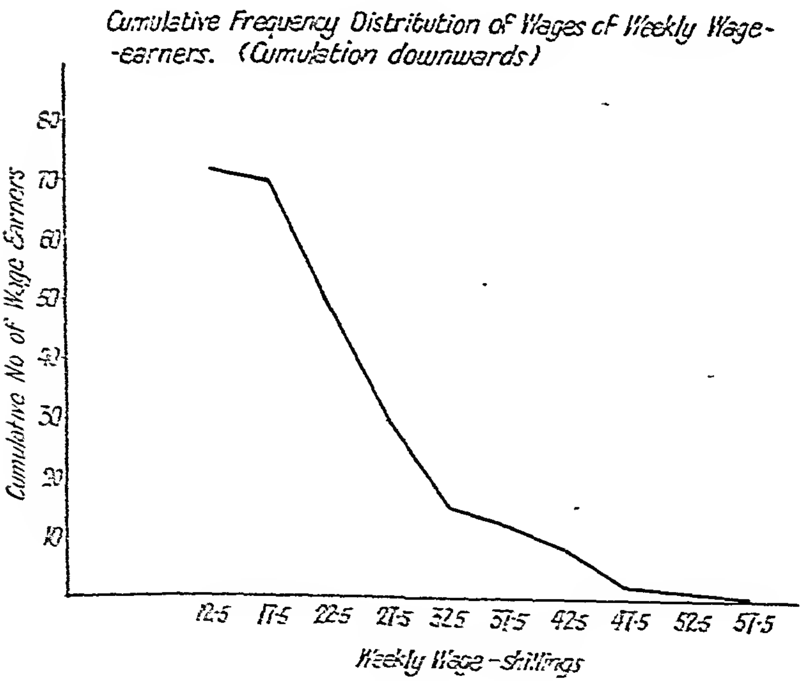

Cumulative Frequency Distribution of Wages of Weekly Wage--earners. (Cumulation downwards)
Cumulative No of Wage Earners
10
20
30
40
50
60
70
80
12·5
17·5
22·5
27·5
32·5
37·5
42·5
47·5
52·5
57·5
Weekly Wage-shillings

FIG. 24

Comparative Frequency Distributions

Two or More Frequency Distributions (Simple or Cumulative) may be Plotted on the Same Chart In order to secure effective com parison it is advisable to reduce the frequencies to a percentage or per mille basis The following is an illustration of the methods available—

TABLE 18

FREQUENCY DISTRIBUTIONS OF ESTATES ACCORDING TO VALUE[1]

VALUE OF ESTATE		STATE B			STATE C		
		No of Estates	Per cent of Total	Cumu lative ditto	No of Estates	Per cent of Total	Cumu lative ditto
(1)		(2)	(3)	(4)	(5)	(6)	(7)
£000)	and under (£000)						
0	2	1	0 5	0 5	2	0 5	0 0
2	4	5	2 5	3 0	9	2 1	2 6
4	6	16	7 9	10 9	20	4 7	7 3
6	8	25	12 3	23 2	34	8 0	15 3
8	10	12	5 9	29 1	44	10 4	25 7
10	12	9	4 4	33 5	56	13 2	38 9
12	14	17	8 4	41 9	68	16 1	55 0
14	16	35	17 2	59 1	70	16 6	71 6
16	18	35	17 2	76 3	62	14 7	86 3
18	20	18	8 9	85 2	50	11 8	98 1
20	—	30	14 8	100 0	8	1 9	100
TOTALS		203	100 0		423	100 0	

STATE B

a = £13 750 M = £14 940 Z is indeterminate
D_1 = £5 790 Q_1 = £8 625 Q_3 = £17 840 D_9 = £20 650
η = £4 625 σ = £5 415 QD = £4 610
CV = 39 38 per cent

STATE C

a = £12 970 M = £13 370 Z = £14 400
D_1 = £6 630 Q_1 = £9 850 Q_3 = £16 460 D_9 = £18 630
η = £3 600 σ = £4 400 QD = £3 395
CV = 33 96 per cent j_1 = 0 45

See Figs 25 and 26

The Array

Fig 27 exhibits the results of arraying the data in Table 12 There are 72 vertical lines, one for each wage earner and proportional

[1] Society of Incorporated Accountants, May, 1931

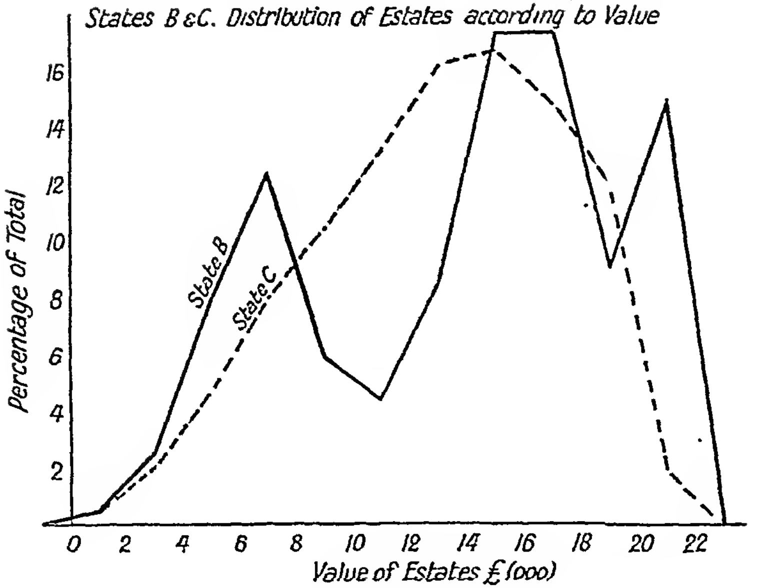

States B & C. Distribution of Estates according to Value
Percentage of Total
16
14
12
10
8
6
4
2
State B
State C
0
2
4
6
8
10
12
14
16
18
20
22
Value of Estates £ (000)

FIG. 25

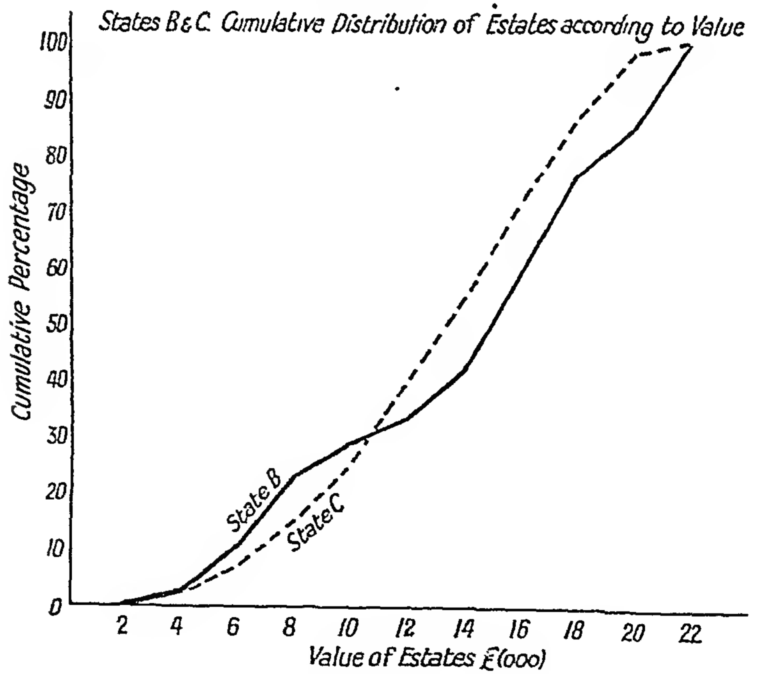

States B & C Cumulative Distribution of Estates according to Value
Cumulative Percentage
100
90
80
70
60
50
40
30
20
10
0
State B
State C
2
4
6
8
10
12
14
16
18
20
22
Value of Estates £(000)

FIG. 26

to the magnitude of his wage The curve is in the form of an ogive, very similar to the curves on page 71, except that it lies on its side.

The Median and Quartiles[1] have been marked on this figure

The Grouped Array.

Anticipating the results of Chapter XI, page 105 on the subject of Quartiles, Octiles, and Deciles, we may test the result of tabulating

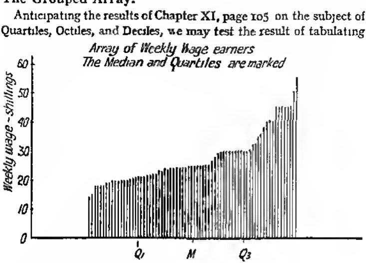


FIG 27

the 72 items in 4 blocks of 18 or 8 blocks of 9 or 10 blocks of 7 or 8 items Let us choose the last, as the most difficult

TABLE 19

GROUPED ARRAY OF WEEKLY WAGE EARNERS ACCORDING TO WAGES

Group No (1)	Rank of Items Included (2)	Average Wage of Earners in Col (2) (3)
		s
1	1 7	17 1
2	8 14	19 6
3	15–22	21 1
4	23–29	23 1
5	30–36	24 3
6	37 43	25 4
7	44 50	29 6
8	51–58	35 1
9	59–65	39 ·
10	66–72	47 1

[1] For an explanation of these terms see Chapters X and XI

Since 72 is not exactly divisible by 10, it is necessary to adopt some convention regarding the odd items. Accordingly, groups Nos. 3 and 8 contain 8 items each and all the rest 7. This method is not often employed, but it is occasionally useful, e.g. to find the average incomes of the richest and the poorest tenths of a population.

Statistical Population.

The term population denotes the totality of objects of which a given group forms part. A population may be determinate in extent (e.g. total make of product A by factory X over a standard period), but more often it is indeterminate (e.g. total world output of product B—past, present, and future).

The population provides a norm with which particular sample values may be compared. The process of smoothing[1] amounts to an attempt to reconstruct population values.

The notion of Population is latent throughout Statistical Theory. Most analysis resolves itself into tests whether a given sample can reasonably be regarded as a sample from a population with given characteristics, or not.

[1] See page 68.

CHAPTER X

STATISTICAL AVERAGES

WHILST the Frequency Distribution achieves a high degree of compression in an otherwise unmanageable mass of raw material it frequently fails to carry the process far enough A table containing ten to thirty entries is still too diffuse to be readily grasped and more powerful methods must be employed

This and succeeding chapters will show how a Statistical Group can be concisely described by reference to three quantities only viz —

1 The Average (or Mean) which indicates the size of the representative item of the group

2 Dispersion which measures the extent to which the items comprised in the group vary in size

3 Skewness which is a measure of the tendency of the group towards asymmetry (or lop-sidedness)

The Average (or Mean)

A statistical group must be composed of homogeneous items i e items alike in relevant aspects for if the items were not alike there would be no justification for grouping them together Upon these grounds it is possible without doing violence to the facts to choose from the group a typical item to represent that group Such a typical item can then be substituted for the individual items in further calculation The object of an average is to describe the group it represents and to afford a basis of comparison with other groups

There are various lines of thought leading up to the choice of type and each involves its own kind of average Each kind has its special advantages and drawbacks which will be considered later

Kinds of Average

There are four kinds of Average (or Mean) in common use viz —

1 The Arithmetic Average (Arithmetic Mean)
2 The Geometric Average (Geometric Mean)

3. The Median.
4. The Mode.

In addition, there are other forms of average such as the Harmonic Mean and the Quadratic Mean, with which we shall not concern ourselves.

Arithmetic Average

The Arithmetic Average (Arithmetic Mean) is the sum of the values of the items concerned, divided by their number, i.e. the average of common speech.

We may distinguish—

1. The Simple Average, in which each item is counted once only and
2. The Weighted Average, in which each item is assigned a weight proportional to its importance in the system.

Simple Arithmetic Average.

Each item is counted once only. If two or more items are identical in size they must be repeated accordingly.

On 1st September, 1931, dealings in 5 per cent *War Loan*, 1929–47, took place at the following prices according to the *Stock Exchange Official List*.

$100\frac{1}{16}$ $\frac{3}{8}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{1}{16}$ $\frac{1}{16}$ $\frac{5}{16}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{3}{8}$ $\frac{1}{16}$ $\frac{3}{8}$ $\frac{9}{32}$ $\frac{11}{32}$ $\frac{1}{16}$ $\frac{1}{8}$ $\frac{1}{8}$
$\frac{1}{8}$ $\frac{1}{16}$ $\frac{1}{16}$ 100 100 $99\frac{7}{8}$ $100\frac{9}{32}$ $\frac{9}{32}$ $\frac{9}{32}$ $\frac{11}{32}$ $\frac{7}{32}$ $\frac{7}{32}$ $\frac{1}{32}$ $\frac{11}{32}$

No. of markings	31
Total	$3,106\frac{1}{4}$

$$\text{Average price} = \frac{3106\frac{1}{4}}{31} = 100\tfrac{13}{64} \text{ nearly}$$

Assuming, for the sake of argument, that the *Stock Exchange List* contained a complete record of bargains in *War Loan* on that day, and that all bargains were of equal importance, the figure $100\frac{13}{64}$ represents the average or typical price at which stock changed hands that day.

Composite Averages.

Simple Averages may be combined to form composite averages.

The following tables are based on the monthly returns of the ten *London Clearing Banks*.

TABLE 20

LONDON CLEARING BANKS—MONTHLY RETURN OF ADVANCES 1924[1]

Date	Advances (i e Loans and Overdrafts)	Date	Advances (i e Loans and Overdrafts)
(1)	(2)	(3)	(2)
	(£000 000)		(£000 000)
January	744	July	775
February	755	August	773
March	772	September	778
April	772	October	787
May	772	November	790
June	778	December	795
		TOTAL	9 291

Average for year $= \frac{9291}{12} =$ £774 25 million

[1] *Committee on Finance and Industry Report* 1931 (Cmd 3897) p 286

TABLE 21

LONDON CLEARING BANKS—ADVANCES—SUMMARY OF MONTHLY AVERAGES 1924–30[1]

Year	Advances (Monthly Averages)
(1)	(2)
	(£000 000)
1924	774 2
1925	821 6
1926	858 0
1927	899 2
1928	923 6
1929	964 0
1930	933 4
AVERAGE 1924–30	882 0

[1] *Ibid* p 296

Simple Arithmetic Average—Algebraic Treatment.

The variate is usually denoted by x, and its individual values by

$$x_1, x_2 \quad x_n$$

The number of items is denoted by n

The arithmetic average is usually denoted by $\bar{x}$ a, or m

The formula for the simple arithmetic average is

$$a = \bar{x} = \frac{1}{n}\left\{x_1 + x_2 \; . \; . \; . + x_n\right\} \quad . \quad . \quad . \quad . \quad (1)$$

In the example of page 79 above

$$a = \tfrac{1}{31}\left\{100\tfrac{5}{16} + 100\tfrac{3}{8} + \; . \; . \; . + 100\tfrac{11}{32}\right\}$$

Formula No. (1) may be abbreviated in the form

$$a = \frac{\Sigma(x)}{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Where Σ (Sigma) is a sign of summation and Σ (x) denotes the sum of all quantities like x.

1. *The value of the arithmetic average is independent of any change in the origin of measurement.*

Proof—

$$a = \frac{\Sigma(x)}{n}$$

$$a - c = \frac{\Sigma(x) - nc}{n} = \frac{\Sigma(x - c)}{n} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Example—

Let the successive values of x be—

13 7 92 64 48 15 19 22 . . . ft.

$$a = \frac{\Sigma(x)}{n} = \frac{280}{8} = 35 \text{ ft.}$$

Changing the origin of measurement so that we henceforth measure the distance from 30 ft., i.e. putting $c = 30$, the successive values become—

-17 -23 62 34 18 -15 -11 -8 ft.

$$\frac{\Sigma(x - c)}{n} = \frac{40}{8} = 5 \text{ ft.} = a - c$$

2. *The value of the arithmetic average is independent of the unit of measurement.*

Proof—

$$a = \frac{\Sigma(x)}{n}$$

$$ra = \frac{\Sigma(rx)}{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Example—

Change the unit of measurement by converting to inches

$$\frac{\Sigma(x)}{n} = \frac{3360}{8} = 420 \text{ in}$$

$$\frac{\Sigma(x-c)}{n} = \frac{480}{8} = 60 \text{ in}$$

3 *The algebraic sum of the deviations of the individual items from the arithmetic average is zero*

Proof—

$$\frac{\Sigma(x)}{n} = a$$

$$\frac{\Sigma(x) - na}{n} = 0 = \frac{\Sigma(x-a)}{n} \qquad (5)$$

Example—

The respective deviations of x from a are

$-22 \quad -28 \quad 57 \quad 20 \quad 13 \quad -20 \quad -16 \quad 13$

And their sum is zero

The quantity $(x-a)$ is known as a deviation It is often denoted by d where d may represent either a positive or a negative quantity

Weighted Arithmetic Average.

So far it has been assumed that every item is of equal importance and is to be counted only once With a Weighted Average such is not the case The items vary in importance, and in calculating the average each must be multiplied by a weight proportional to the extent of its importance

Actual and Estimated Weights

The weights should be based upon actual figures where these are forthcoming otherwise they must be estimated from the best data available The effect of using approximate or estimated weights is brought out below

Consider the following table relating to wages paid by a manufacturing establishment [1]

[1] The figures are imaginary

TABLE 22

WEEKLY WAGES PAID IN THE X ESTABLISHMENT
(Weighted Averages)

Weekly Wage (1)	Actual No. of Operatives (2)	Product of Cols. (1) & (2) (3)	Approximate Weights (A) (4)	Product of Cols. (1) & (4) (5)	Approximate Weights (B) (6)	Product of Cols. (1) & (6) (7)
s.						
35	135	4,725	7	245	1	35
42	471	19,782	24	1,008	5	210
50	628	31,400	31	1,550	6	300
54	514	27,756	26	1,404	5	270
60	97	5,820	5	300	1	60
TOTALS .	1,845	89,483	93	4,507	18	875

Weighted Average in—

Actual weights $= \frac{89483}{1845} = 48 \cdot 50s.$

Approximate weights A $= \frac{4507}{93} = 48 \cdot 46s.$

" " B $= \frac{875}{18} = 48 \cdot 61s.$

Column (2) shows the number of operatives employed at each wage. These figures are evidently *actual* weights. Column (4) is found by dividing column (2) by 20 to the nearest integer, and column (6) is found by dividing column (2) by 100 to the nearest integer. The figures in these columns are *approximate* weights: they might be the results of estimates if the actual numbers were unknown.

In either case the error involved by the use of approximate weights is relatively small and can for many practical purposes be neglected.

In the above example *actual* or *estimated* weights are applied to *actual* figures. Let us now consider the application of *estimated* weights to *estimated* figures.

The *Ministry of Labour Cost of Living Index*[1] is a weighted average of percentage increases in five groups of items entering into the budget of an urban working class family. Both the percentage increases and the weights are estimated. The following statement gives the make-up of the figure as at 1st May, 1931.

[1] For full particulars, see Chapter XXI.

2 Write down column (2) in the form shown Then ξ measures the wage in terms of 5s units reckoned from the point in question

3 Find the value of the average in units by the rule of Table 24 and convert the result into shillings as shown

TABLE 25

SHORT-CUT METHOD OF FINDING ARITHMETIC AVERAGE WAGE OF WEEKLY WAGE EARNERS

Central Wage ($= x$)	Distance from Assumed Average ($= \xi$)	Frequency ($= f$)	Product of Cols (2) and (3) ($= f\xi$)
(1)	(2)	(3)	(4)
s	Units		
15	− 2	2	− 4
20	1	22	− 22
25	0	19	0
30	+ 1	14	+ 14
35	+ 2	3	+ 6
40	+ 3	4	+ 12
45	+ 4	6	+ 24
50	+ 5	1	+ 5
55	+ 6	1	+ 6
		72 ($= \Sigma f$)	+ 41 ($= \Sigma f\xi$)

Then the average wage $= 5 + \frac{41}{72}$ units

$$= 25 + \frac{41 \times 5}{72} = 27\ 85 \text{ shillings}$$

agreeing exactly with the answer obtained on page 83

The saving in labour is obvious The answer obtained by this method can be checked independently by re computing from another arbitrary point The student should test the result of calculating from 20s and 30s as origin taking care of the algebraic sign in the latter case

Algebraic Proof

By the rule for weighted averages we have—

$$\frac{\Sigma(f\xi)}{\Sigma(f)} = \frac{72}{41} \text{ units}$$

Columns (1) and (2) are related by the equation

$$x = 5\xi + 25$$

This equation must be true of the average. Consequently,

$$a = 5 \times \frac{72}{41} + 25 \text{ shillings}$$

The general expression is

$$x = r\xi + t \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Where r represents the width of the group interval and t the trial average.

The calculation of the arithmetic average of a frequency distribution involves a slight error due to the assumption that the values in each group are all equal to the central value. The true average calculated from the individual values given in Table 12 is 27·69s., the difference between the true and the computed value being 0·16s., involving an error of 0·6 per cent.

Unequal Group Intervals.

By extension of this principle we may calculate the average when the group intervals are irregular in width. Let us find the average age of the gainfully occupied male population shown in Table 16, checking the result.

TABLE 26

CALCULATION OF AVERAGE AGE OF GAINFULLY OCCUPIED MALE POPULATION OF GREAT BRITAIN, 1921

Central Age (x)	Frequency (f)	ξ	$f\xi$	Check	
				ζ	$f\zeta$
(1)	(2)	(3)	(4)	(5)	(6)
	(000's)				
13	44	−54	− 2,376	−34	− 1,496
15	532	−50	−26,600	−30	−15,960
17	725	−46	−33,350	−26	−18,850
19	739	−42	−31,038	−22	−16,258
22·5	1,601	−35	−56,035	−15	−24,015
30	2,887	−20	−57,740	0	0
40	2,731	0	0	+ 20	+ 54,620
50	2,318	+ 20	+ 46,360	+ 40	+ 92,720
60	1,429	+ 40	+ 57,160	+ 60	+ 85,740
67·5	404	+ 55	+ 22,220	+ 75	+ 30,300
72·5	246	+ 65	+ 15,990	+ 85	+ 20,910
TOTALS	13,656		−65,409		+ 207,711

Then $a = 80 - \frac{65\,409}{13\,656}$ units

$= 40 - \frac{65\,409}{13\,656 \times 2} = 37\ 61$ years

Check $a = 60 + \frac{207\,711}{13\,656}$ units

$= 30 + \frac{207\,711}{13\,656 \times 2} = 37\ 61$ years

In order to secure round numbers in columns (3) and (5) the unit was taken as 0 5 years The arbitrary origin chosen for the first trial (40 years = 80 units) proved higher than the true value so that the correction was a minus quantity

Advantages and Disadvantages of the Arithmetic Mean as a Type

Advantages—

1 *It is easy to* understand *and calculate*
2 It utilizes *all* the data in the group
3 It is determinate
4 It is suitable for arithmetic and algebraic manipulation

Disadvantages—

1 It may give too much weight to extreme (and therefore abnormal) items

2 It may locate the type at a point at which few (or none) of the actual observations lie

The arithmetic average is the most useful general purpose average It should always be employed unless there are special reasons for choosing some other type

Geometric Average (Geometric Mean)

The geometric mean is the *n*th root of the product of the *n* quantities comprised in the group

Symbolically

$$g = \sqrt[n]{x_1 \times x_2 \times \ldots x_n}$$
$$= \sqrt[n]{\Pi x} \qquad (11)$$

where Π (Pi) is a sign of multiplication

The weighted geometric mean is symbolized by

$$g = \sqrt[n]{x_1^{w_1} \times x_2^{w_2} \times \ldots x_n^{w_n}} \qquad (12)$$

where the w's represent the weights and $\Sigma(w) = n$.

In practice the geometric mean is calculated by means of logarithms, i.e.

$$\log g = \frac{1}{n}\{\log x_1 + \log x_2 \ldots + \log x_n\}$$

$$= \frac{1}{n}\Sigma(\log x) \quad . \quad . \quad . \quad . \quad . \quad . \qquad (13)$$

The geometric average is less than the corresponding arithmetic average unless all the quantities are equal.

The geometric average is chiefly employed in connection with index numbers.[1] Index numbers are ratios, and the geometric

TABLE 27

BOARD OF TRADE INDEX OF WHOLESALE PRICES, JULY, 1931.[2]
CALCULATION OF FINAL INDICES BASED ON GEOMETRIC MEAN

Group (1)	No. of Items Included (2)	Price Index No. (1924 = 100) (3)	Log (3) (4)	Col. (2) × Col. (4) (5)	Col. (2) × Col. (3) (6)
I Cereals	17	53·1	1·7251	29·3267	902·7
II Meat and Fish	17	74·7	1·8733	31·8461	1269·9
III Other Foods	19	71·6	1·8549	35·2431	1360·4
TOTAL	53			96·4159	3533·0
IV. Iron and Steel	24	72·1	1·8579	44·5896	1730·4
V. Coal	10	68·8	1·8376	18·3760	688·0
VI. Other Metals and Minerals	10	62·2	1·7938	17·9380	622·0
VII. Cotton	16	42·3	1·6263	26·0208	676·8
VIII. Wool	9	43·8	1·6415	14·7735	394·2
IX. Other Textiles	6	46·3	1·6656	9·9936	277·8
X. Miscellaneous	22	67·5	1·8293	40·2446	1485·0
TOTAL	97			171·9361	5874·2
GRAND TOTAL	150			268·3520	9407·2

[1] See Chapter XVI.

[2] *Board of Trade Journal*, 13th August, 1931 (p. 183).

average has certain properties that make it especially useful when dealing with relative as contrasted with absolute numbers

The *Board of Trade Index Number of Wholesale Prices* utilizes the geometric mean

In the Table 27 (p 67) column (2) shows the w s (No of items in this case) column (3) the x s and column (4) the log x s Columns (5) and (6) show the results of applying the w s to the log x s and the x s respectively

The Table below shows the computation of the final indices for July 1931 The arithmetic means are also shown for purposes of comparison

	Geometric Mean	Arithmetic Mean (for comparison)
I III	$\log g \quad \frac{96\ 4159}{53} = 1\ 8191$ $g = 65\ 9$	$a = \frac{3533\ 0}{53} = 66\ 7$
IV–X	$\log g \quad \frac{171\ 9361}{97} \quad 1\ 7725$ $g \quad 59\ 2$	$a \quad \frac{5874\ 2}{97} = 60\ 6$
I X	$\log g = \frac{268\ 3520}{150} = 1\ 7890$ $g \quad 61\ 5$	$a = \frac{9407\ 2}{150} = 62\ 7$

Since this section was originally written the base year of the *Board of Trade* Index has been changed from 1924 to 1930 This does not affect the principle

Advantages and Disadvantages of the Geometric Mean as a Type

Advantages—

1 It utilizes all the data in the group

2 It is determinate provided that all the quantities are greater than zero

3 It is suitable for arithmetic and algebraic manipulation

4 It attaches less weight to large items than does the arithmetic mean

5 It is especially suitable for ratios

Disadvantages—

1 It cannot be used when any of the quantities are zero or negative

2. It is less easy to understand and calculate than the arithmetic mean.

3. It may locate the type at a point at which few (or none) of the actual observations lie.

The Median

The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater, and the other all values less than the median.

The median depends upon the rank or position of the item concerned; hence it is necessary to array[1] the items in order to find it.

Required to find the median of the *Stock Exchange* transactions given on page 77. Arraying the items in order of magnitude we have—

$99\frac{7}{8}$ 100 100 $100\frac{1}{32}$ $\frac{1}{16}$ $\frac{1}{16}$ $\frac{1}{16}$ $\frac{3}{32}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{3}{16}$ $\frac{3}{16}$ $\frac{3}{16}$

$\frac{7}{32}$ $\frac{7}{32}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{9}{32}$ $\frac{9}{32}$ $\frac{9}{32}$ $\frac{5}{16}$ $\frac{5}{16}$ $\frac{5}{16}$ $\frac{11}{32}$ $\frac{11}{32}$ $\frac{11}{32}$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{3}{8}$

There are thirty-one items, and the median is evidently the size of item No. 16, viz. $100\frac{7}{32}$. Note that in this case the median differs only slightly from the arithmetic average: $100\frac{13}{64}$.

When the number of items is even, the median is intermediate between the sizes of the two middle items. In such case it is usual to take the average of the two.

If the values of the items are erratic, it is advisable to use the Extended Median found by averaging (say) the five middle items.

The median wage of the wage-earners shown in Table 12 lies between items 36 and 37, i.e. at 25s.

Median of Continuous Frequency Distribution

In this case the median cannot be found *directly* without recourse to the original data. It may, however, be *estimated* with sufficient accuracy by interpolation.

Consider the distribution of estates according to value as given in Table 18 (p. 72). The figures are cumulated in Table 28.

[1] See Chapter IX, p. 74.

TABLE 28

CUMULATIVE FREQUENCY DISTRIBUTION OF ESTATES ACCORDING TO VALUE

Value Not Exceeding	Number of Estates	
	State B	State C
(1)	(2)	(3)
(£000 s)		
2	1	2
4	6	11
6	22	31
8	47	65
10	59	109
12	68	165
14	85	233
16	120	303
18	155	365
20	173	415
—	203	423

STATE B

Assuming for this purpose that the items are distributed evenly within their respective grades we have the following scheme—

Rank of Item	Corresponding Value	Rank of Item	Corresponding Value
	£		£
85	—	103	—
—	14 000	—	15 029
86	—	—	—
—	14 057	—	—
87	—	119	—
—	14 114	—	15 943
—	—	120	—
—	—	—	16 000
101	—	121	—
—	14 914 }	—	16 057
102	— } 14 940 (say)		
—	14 971 }		

On this part of the range an advance of 35 ranks corresponds with an advance of £2 000 viz £57 14s per rank

Therefore an advance of 16½ ranks corresponds with an advance of $\frac{16\ 5}{35} \times$ £2000

Therefore the median (size of the middle item) is

$$14{,}000 + \frac{16\cdot5}{35} \times 2000 = £14{,}940 \text{ (say)}$$

Symbolically, this may be written—

$$M = l_1 + \frac{\frac{n}{2} - r_1}{r_2 - r_1} \times (l_2 - l_1) \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Where M represents the median[1]

n	,,	,, No. of items in the group
l_2 and l_1	,,	,, limits of the class containing the median
r_2 and r_1	,,	,, ranks of the items just below those limits

STATE C.

Applying this formula we have

$$M = 12{,}000 + \frac{211\cdot5 - 165}{233 - 165} \times 2000$$
$$= £13{,}370^{2}$$

As a check upon this calculation, we can find the median by cumulating upwards instead of downwards. The student should verify that upon this basis

$$M = 14{,}000 - \frac{211\cdot5 - 190}{258 - 190} \times 2000$$
$$= £13{,}370 \text{ (as before)}$$

The median wage of the weekly wage-earners recorded in Table 13 is given by

$$M = 22\cdot5 + \frac{36 - 24}{43 - 24} \times 5$$
$$= 25\cdot66\text{s.}$$

The corresponding figure by the direct method is 25s.[3]

If the distribution is discrete, the value of the median may be found by the ordinary rule without interpolation.

[1] Remember that the median is the *size* of the middle item, *not its rank*.
[2] Interpolation only gives an *approximate* value.
[3] See p. 89.

Graphic Method

A graphic method of finding the median is given on page 106

Advantages and Disadvantages of the Median

Advantages—

1 If found directly it represents an actual item

2 It is easy to understand

3 It eliminates the effect of extreme (and therefore abnormal) items

4 It can be utilized for incommensurable items

5 Only the values of the middle items need be known

Disadvantages—

1 If the distribution is irregular the indication of the median may be indefinite

2 It cannot be located with precision when the items are grouped

3 It is not suitable for arithmetic or algebraic manipulation

The Mode

The Mode is the size of the variable that occurs most frequently or the position of greatest density In a smoothed histogram it is represented by the position of the maximum ordinate [1]

Local inquiries into wages frequently require the current wage or the usual wage This must be understood to refer to the modal wage

In connection with their *Cost of Living Statistics* the *Ministry of Labour* make inquiries as to predominant rents in the districts concerned

Inquiries as to modal wages rents prices etc can frequently be answered off hand by persons of experience whilst inquiries as to average quantities would involve a considerable amount of labour in collection

The organizer of mass production will adjust his standards according to the most common demand and by so doing will satisfy the requirements of the bulk of his customers Were he to aim at average standards he might satisfy no one

[1] *The mode is the value measured along the x axis*

Methods of Determining the Mode.

With a smoothed histogram the mode can be located exactly. It is the point on the x-axis corresponding to the peak of the curve.

If it is inconvenient or impossible to construct a smoothed histogram, the mode may frequently be located by grouping.

Required to find the modal class in the following table by the group method and to determine the mode within that class by formula.[1]

TABLE 29

DETERMINATION OF MODE BY GROUPING

Variable	Frequency					
1–2	1	5		12		
2–3	4		11		19	
3–4	7	15				26
4–5	8		19	32		
5–6	11	24			**37**	
6–7	13		**26**			**40**
7–8	13	**27**		**37**		
8–9	14		24		36	
9–10	10	22				30
10–11	12		20	26		
11–12	8	14			17	
12–13	6		9			11
13–14	3	5				
14–15	2					

The frequencies are grouped in two's, then in three's, and the maximum of each column is indicated in heavy type. It will be seen that the

7–8 group occurs in the maximum 5 times
8–9 ,, ,, ,, 3 ,,
6–7 ,, ,, ,, 3 ,,

[1] Society of Incorporated Accountants, May, 1931.

Evidently the modal group is the 7–8 group

To find the mode more precisely the following formula may be used

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 \quad f_0 - f_2}(l_2 - l_1) \quad . \qquad (15)^1$$

Where Z represents the mode

l_1 and l_2 „ limits of the modal group

$f_0 f_1$ and f_2 „ . frequencies in the three groups of which the modal group forms the centre

Applying the formula we have

$$Z = 7 + \frac{13 - 13}{26 \quad 13 - 14}(8 \quad 7)$$
$$= 7$$

Provided the figures run fairly regularly the mode can be calculated directly from the grouped frequency distribution without any preliminaries

Thus in Table 13 the principal mode is given by

$$Z_1 = 17\ 5 + \frac{22 - 2}{44 \quad 2 - 19}(22\ 5 - 17\ 5)$$
$$= 21\ 855$$

There is a secondary mode given by

$$Z_2 = 42\ 5 + \frac{6 \quad 4}{12 - 4 \quad 1}(47\ 5 - 42\ 5)$$
$$= 43\ 935$$

These values correspond almost exactly with the values given by the smoothed curve

Advantages and Disadvantages of the Mode

Advantages—

1 It is easily understood

2 It eliminates the effect of extreme (and therefore abnormal) items

3 Only the values of the middle items need be known

[1] This formula (which is recommended by Professor Bowley) has been substituted for the customary (and less accurate) formula viz —

$$Z = l_1 + \frac{f_2}{f_2 + f_0}(l_2 \quad l_1)$$

Disadvantages—

1. It is frequently ill-defined.
2. It is difficult to locate exactly.
3. It is unsuitable for arithmetic and algebraic manipulation.

Frequency Distribution of Wages of Weekly Wage-earners
(Showing arithmetic average, median & mode)

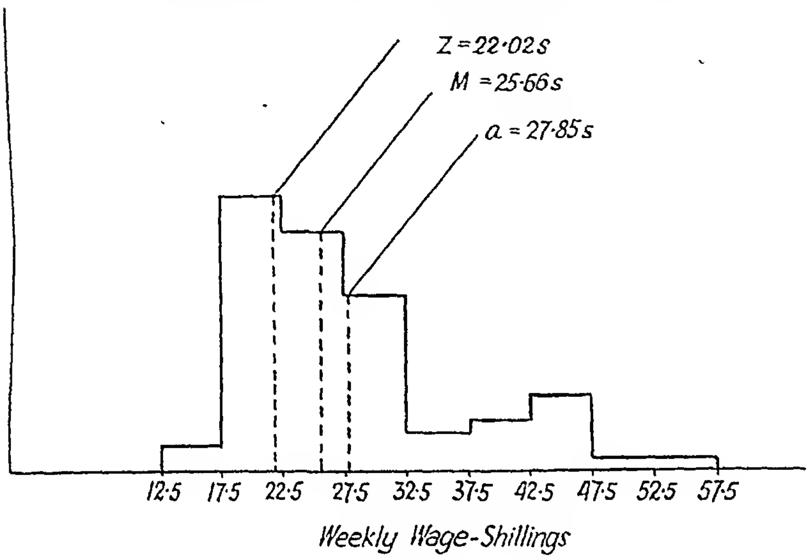


FIG. 28

Representation of Averages on a Histogram.

In the above diagram Fig. 19 has been redrawn so as to indicate the location of the arithmetic average, median, and mode.

Note that the mode does not fall in the middle of the largest group. This is due to groups 3 and 4, which pull the mode over to the right.

Had the group been symmetrical,[1] the three averages would have coincided. Owing, however, to the considerable amount of skewness[1] present, the arithmetic average and the median have been pulled away from the mode, the former more than the latter.

It will be realized that the precise location of these averages on the graph depends somewhat upon the grouping of the individual

[1] See Chapter XII.

items In this respect the mode and the median are more sensitive than the arithmetic mean, and this constitutes an important argument in favour of the latter

Typical and Descriptive Averages.

Although the calculation of an average from a given set of numerical data is always arithmetically possible, it does not follow that the calculation is statistically significant nor that it is worth performing If the average, as calculated falls near a point round which the data tend to cluster, then it may reasonably be assumed to represent the group In this case it is called a "Typical Average" If, however, the data cluster round several points or fail to cluster at all, it is a "Descriptive Average" and possesses merely arithmetical significance To take an exaggerated instance the "average' income of a group consisting of 99 persons with £100 apiece and one person with £40,100 would be £500 This proposition would be statistically meaningless unless of course, it were proposed to expropriate the single individual in question and divide the proceeds equally all round

Choice of Average

In the absence of special circumstances it is usual to employ the arithmetic mean for general purposes and the geometric mean when dealing with ratios or index numbers The reason is that these forms of average utilize *all* the information available, and consequently tend to be less erratic, i e less sensitive to small changes in individual values The comparative ease with which they can be manipulated also constitutes a strong argument in their favour The median may be useful in cases in which the extreme values are ill-defined The mode is seldom used in elementary work owing to the difficulty in ascertaining it with precision

Standardized Death Rates

In order to effect valid comparisons between death rates of different localities it is necessary to eliminate differences between age and sex constitution of their populations First we must break up the population into groups according to the incidence of death rates, then apply the local death rates to a standard population

(e.g. the population of the country at large). Here is a simple illustration—

Age—Years	0–5	5–15	15–65	65–	Total
Standard Population—					
Age Constitution . .	75	250	600	75	1,000
Death Rate per 1,000	25	5	7	65	—
Local Population—					
Age Constitution . .	50	260	630	60	1,000
Death Rate per 1,000 .	30	6	8	70	—

Death Rate—Standard Population =

$\frac{1}{1000}(75 \times 25 + 250 \times 5 + 600 \times 7 + 75 \times 65) = 12{\cdot}2$ per 1,000.

Crude Death Rate—Local Population =

$\frac{1}{1000}(50 \times 30 + 260 \times 6 + 630 \times 8 + 60 \times 70) = 12{\cdot}3$ per 1,000

Standardized Death Rate—Local Population =

$\frac{1}{1000}(75 \times 30 + 250 \times 6 + 600 \times 8 + 75 \times 70) = 13{\cdot}8$ per 1,000.

Had we relied on the indications of the crude death rate we should have noticed nothing remarkable. The standardized death rate shows, however, that the local death rate is in reality the higher.

This method is of general application. For instance, we may standardize unemployment rates by eliminating differences in occupations of unemployed persons.

CHAPTER XI

DISPERSION

THE average is a typical member of the group and represents that group For the statistician that is not enough he wants to know in addition *to what extent* the average is typ cal in other words how the items comprised in the group vary in size

Dispersion is a Measure of the extent to which the Individual Items vary What we are concerned with is not the absolute size of the items but the magnitude of their deviations from their type For this reason measures of dispersion may be referred to as averages of the second order

Measures of Dispersion

There are four measures of dispersion in common use—

1 The Range
2 The Mean Deviation (Average Deviation)
3 The Standard Deviation
4 The Quartile Deviation

THE RANGE

The Range is represented by the difference between the sizes of the largest and the smallest item

In the Stock Exchange example on page 77 the smallest item is $99\frac{7}{8}$ and the largest $100\frac{3}{8}$ The Range is therefore $\frac{1}{2}$

The Range is not a satisfactory measure of dispersion since it depends entirely upon the sizes of extreme (and possibly abnormal) items [1]

THE MEAN DEVIATION (AVERAGE DEVIATION)

As already explained the (algebraic) difference between an individual item and its type is called a deviation (usually symbolized by *d*)

Deviations can be reckoned from the arithmetic mean the median or the mode

[1] Calculations depending upon the *average* ranges of a number of samples do lead to useful results The subject is however too complex to pursue here

$d_a = x - a$ deviation from the arithmetic mean

$d_m = x - M$ „ „ median

$d_z = x - Z$ „ „ mode

The Mean Deviation is the arithmetic average of the deviations of the group (all taken as positive),[1] i.e. their sum divided by their number.

The larger the differences between the items and the type, the larger will be the mean deviation; if all the items are identical the mean deviation will be zero.

Technically, it is best to calculate the mean deviation from the median. Frequently, however, it is calculated from the arithmetic mean. Calculation from the Mode is unusual

Calculation from the Median.

In the Stock Exchange example on page 77 the median is $100\frac{5}{32}$. The respective deviations $(+)$ are therefore—

11 7 7 6 5 5 5 4 3 3 3 1 1
1 0 0 1 1 1 2 2 2 3 3 3
4 4 4 5 5 5: 32nds.

$$\text{Sum of the above} = \frac{107}{32}$$

$$\text{No.}\quad „ \quad „ \quad = 31$$

$$\text{Mean deviation} = \frac{107}{32 \times 31} \doteq 0{\cdot}1079 = \frac{7}{64}\ \text{(approximately)}.$$

In practice, we do not go to the trouble of finding the individual deviations, but rely on the fact that η (the Mean deviation) is equivalent to $\frac{1}{n}$th of the difference between the values above the median and the values below. In this example the sum of the 15 items above the median is $1504\frac{22}{32}$ and the sum of the 15 items below is $1501\frac{11}{32}$. The difference is $\frac{107}{32}$, precisely as before.

Proof (for odd values of n)

Write x_1 for the $\frac{n-1}{2}$ values of $x > M$ and x_2 for the $\frac{n-1}{2}$ values $< M$.

Then average deviation $= \eta$

[1] It is necessary to eliminate the sign. The algebraic sum of the deviation from a is zero, and from M or Z nearly zero. (See p. 80.)

$$= \frac{1}{n}\left\{\Sigma(x_1 - M) + \Sigma(M - x_2)\right\}$$

$$= \frac{1}{n}\left\{\Sigma(x_1) - \Sigma(x_2)\right\}. \tag{1}$$

since the rest cancels out

There is a similar proof for even values of n

Calculation from the Arithmetic Mean.

The arithmetic mean is $100\frac{13}{64}$ and the respective deviations (+) are—

21 13 13 11 9 9 9 7 5 5 5 1 1
1 1 1 3 3 3 5 5 5 7 7 7
9 9 9 11 11 11 64ths

$$\text{Sum of the above} = \frac{217}{64}$$

$$\text{No} \quad ,, \quad ,, \quad = 31$$

$$\text{Mean deviation} = \frac{217}{64 \times 31} = 0{\cdot}1094 = \frac{7}{64} \text{ (approximately)}$$

The figure, as calculated from the median is slightly the smaller [1]

In practice we proceed as follows—

There are 17 values $> a$ and their sum is $1705\frac{1}{8}$

,, 14 ,, $< a$,, ,, $1401\frac{1}{8}$

Therefore

$$\eta = \tfrac{1}{31}\left\{\left(1705\tfrac{1}{8} - 17 \times 100\tfrac{13}{64}\right) + \left(14 \times 100\tfrac{13}{64} - 1401\tfrac{1}{8}\right)\right\}$$

$$= \frac{217}{64 \times 31} \text{ (as before)}$$

Algebraic Treatment

Symbolically the mean deviation is represented by

$$\eta = \frac{\Sigma f\,|x - M|^*}{\Sigma f} \tag{2}$$

[1] The sum of the deviations (plus) is always *least* when taken from the *median*

* $|x - M|$ means $(x - M)$ taken as positive Some writers use δ_M and δ_a to denote the mean derivation from M and a respectively

Standard Deviation

The Standard Deviation is the square root of (the sum of the squares of the individual deviations from the arithmetic mean, divided by their number).

The process of squaring gets rid of the *minus* signs that are so troublesome with the mean deviation, and it weights up large deviations as against small ones.

Required to find the standard deviation of the first 5 natural numbers.

The arithmetic mean is $\frac{1+2+3+4+5}{5} = 3$

The Standard Deviation is represented by

$$\sigma = \sqrt{\frac{2^2+1^2+0+1^2+2^2}{5}} = \sqrt{2} = 1{\cdot}4142$$

The quantity σ^2 is known as the Variance and is denoted by V.

The Standard Deviation of the *Stock Exchange* prices given on page 77 is found as follows—

Sum of squares of deviations from arithmetic mean given on page 77.

$$21^2 + 2\times 13^2 + 11^2 + 3\times 9^2 + 7^2 + 3\times 5^2$$
$$+ 5\times 1^2 + 3\times 3^2 + 3\times 5^2 + 3\times 7^2 + 3\times 9^2 + 3\times 11^2$$
$$= 2127 \div (64)^2$$

No. of items $= 31$

$$\sigma = \sqrt{\frac{2127}{(64^2)\times 31}} = 0{\cdot}129 = \tfrac{1}{8} \text{ (approximately)}$$

Symbolically, the standard deviation may be written

$$\sigma = \sqrt{\frac{\Sigma(x-a)^2}{n}} = \sqrt{\frac{\Sigma d^2}{n}} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $d = x - a$

or if the quantities are weighted

$$\sigma = \sqrt{\frac{\Sigma f(x-a)^2}{\Sigma(f)}} = \sqrt{\frac{\Sigma(fd^2)}{\Sigma(f)}} \quad . \quad . \quad . \quad (4)$$

Standard Deviation—Frequency Distribution.

Direct Method. This may be illustrated by a simple problem in chances. Six coins are tossed up and the number of heads is noted.

This experiment is repeated 64 times The mathematical expectation of 0 1 2 6 heads respectively is as follows Find the standard deviation

TABLE 30

RESULTS OF TOSSING SIX COINS—EXPERIMENT REPEATED 64 TIMES
(Calculation of Standard Deviation)

No of Heads (x) (1)	Expected Frequency (f) (2)	Deviation from A M (d) (3)	Col (3) Squared (d^2) (4)	Col (2) × Col (4) (fd^2) (5)
0	1	3	9	9
1	6	−2	4	24
2	15	1	1	15
3	20	0	0	0
4	15	+1	1	15
5	6	+2	4	24
6	1	+3	9	9
TOTAL	64			96

$$a = 3 \quad \sigma = \sqrt{\frac{96}{64}} = \sqrt{\frac{3}{2}} = 1\ 225$$

Short cut Method The direct method becomes very laborious when the average does not fall on a round number The following table illustrates a systematic method of calculating the mean and standard deviation by the short cut method

TABLE 31

CALCULATION OF MEAN AND STANDARD DEVIATION OF WAGES OF WAGE EARNERS
(Short-cut Method)

Central Wage (x) (1)	Distance from Assumed Average ($= \xi$) (2)	Frequency ($= f$) (3)	Col (2) × Col (3) ($= f\xi$) (4)	Col (2) × Col (4) ($= f\xi^2$) (5)
5				
15	2	2	4	8
20	1	22	22	22
23	0	19	0	0
30	+1	14	+14	14
35	+2	3	+6	12
40	+3	4	+12	36
45	+4	6	+24	96
50	+5	1	+5	25
55	+6	1	+6	36
	TOTALS	72 ($= \Sigma f$)	+41 ($= \Sigma f\xi$)	249 ($= \Sigma f\xi^2$)

$$a = 5 + \frac{41}{72} \text{ units} = 27{\cdot}85 \text{ shillings}$$

$$\sigma = \sqrt{\frac{249}{72} - \left(\frac{41}{72}\right)^2}$$

$$= \sqrt{\frac{16{,}247}{5184}} \text{ units} = 8{\cdot}85 \text{ shillings}$$

Algebraic Proof of Short-cut Method.

$$\sigma^2 = \frac{\Sigma f(x-a)^2}{n} \text{ where } a \text{ represents the } \textit{true} \text{ average}$$

$$= \frac{\Sigma fx^2 - 2a\Sigma fx + na^2}{n}$$

$$= \frac{\Sigma fx^2}{n} - a^2 \text{ since } \Sigma fx = na$$

In other words, the $(\text{S.D.})^2 = \dfrac{\text{sum of squares of items}}{\text{No. of items}}$ less $(\text{A.M.})^2$

This is true, whatever point be taken as the origin of x, provided that a is also measured from the same origin.

Standard Deviation of a Sum or Difference.

Let $z = x + y$ where x and y are variable quantities.

In order to simplify the calculation, let z, x, and y be measured from their arithmetic means.

$$\frac{\Sigma(z)}{n} = \frac{\Sigma(x \pm y)}{n} = 0 \qquad \text{(5)}$$

$$\text{And} \frac{\Sigma(z^2)}{n} = \frac{\Sigma(x \pm y)^2}{n}$$

$$= \frac{\Sigma(x^2)}{n} + \frac{\Sigma(y^2)}{n} \pm \frac{2\Sigma(xy)}{n} \qquad \text{(6)}$$

If x and y are independent, in the sense that positive values of x are just as likely to be associated with negative values of y as with positive values, and *vice versa*, the quantity $\Sigma(xy)$ will tend to vanish, leaving

$$\sigma_z^2 = \sigma_x^2 + \sigma^2_y$$

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma^2_y} \qquad \text{(7)}$$

This analysis may be extended to include any number of quantities, and differences as well as sums.

In other words the variance[1] of the sum (or differences) of a number of independent variables is equivalent to the sum of their variances

Quartile Deviation and similar Measures

The Median has been defined as that value of the variable that divides the whole group into two equal parts

The Quartiles divide the distribution in the ratios 1 3 and 3 1 Thus the Lower Quartile Q_1 the Med an (M) and the Upper Quartile (Q_3) divide the distribution into four quarters

In the Table of Wages given on page 61

Q_1 lies between Nos 18[2] and 19	and its value is	21s
M	Nos 36 and 37	25s
Q_3	Nos 54 and 55	30s

The Quartile Deviation (Semi Inter Quartile Range) is then defined as

$$QD = \frac{Q_3 \quad Q_1}{2} \qquad 4\ 5s \tag{8}$$

A difficulty occurs when the number of items is not exactly divisible by four The following is the simplest method[3] of avoiding fractions—

$$\text{Assume } Q_1 = \text{the value of item } \frac{n+1}{4}$$

$$M = \frac{n+1}{2}$$

$$Q_3 = \frac{3(n+1)}{4}$$

subject to the convention that we disregard $\frac{1}{4}$ s and average for the $\frac{1}{2}$ s

Thus for $n = 72$ Q_1 is taken as the size of item 18 and Q_3 as the size of item 55

M is the average size of items 36 and 37

For $n =$ 3 Q_1 is the average size of items 18 and 19 Q_3 the average size of items 55 and 56 M the size of item 37

[1] See p 1

[2] There are 18 items below and 54 above the line of division between Nos 18 and 19

[3] This method is only approximate for a more elaborate one consult Bowley s *Elements of Statistics* (pages 106–7) The subject is not really important for when n is small the quartiles have a large standard error (see Chapter XIII) and when it s large the difference between the two methods vanishes

Quintiles, Octiles, Deciles, and Percentiles.

The Quintiles, Octiles, Deciles, and Percentiles are values of the variable that divide the group into fifths, eighths, tenths, and hundredths, respectively.

In the Table of Wages given on page 61

The lowest Quintile (Qn_1) is given by the size of No. 15, i.e. 20s.
,, Octile (O_1) ,, ,, No. 9, i.e. 20s.
,, Decile (D_1) ,, ,, No. 7, i.e. 19s.

Percentiles would not be used for a group of only 72 items.

Deciles and percentiles are commonly used in social inquiries. There were in 1929 about 5,000,000 persons in the United Kingdom with incomes of £160 or more. Within this group the ninety-eighth percentile would correspond to an income of about £2,000, and the first percentile to an income only slightly over the tax exemption limit.

Location of Quartiles, etc.—Continuous Frequency Distribution.

In this case the required figure can be found by interpolation upon the principle explained in Chapter X, page 89, as regards the Median.

Required to find the median, lower and upper quartiles, and lower and upper deciles for the groups of estates given in Table 18 (State B).

Quantity Required	State B ($n = 203$)
D_1	$4 + \dfrac{20{\cdot}3 - 6}{16} \times 2 = £5790$
Q_1	$8 + \dfrac{50{\cdot}75 - 47}{12} \times 2 = £8625$
M	$14 + \dfrac{101{\cdot}5 - 85}{35} \times 2 = £14{,}940$
Q_3	$16 + \dfrac{152{\cdot}25 - 120}{35} \times 2 = £17{,}840$
D_9	$20 + \dfrac{182{\cdot}7 - 173}{30} \times 2 = £20{,}650$

The Quartile Deviation is easiest to calculate, but it is liable to be erratic

Graphical Representation.

In the following diagram Fig 19, relating to wages of wage earners, has been re drawn so as to show the positions of η, σ, and Q D

Here $a = 27$ 85s Since $\eta = 7$ 01s $a + \eta = 34$ 86s and $a - \eta = 20$ 84s

Also $\sigma = 8$ 85s, $a + \sigma = 36$ 70s, and $a - \sigma = 19$ 00s

If desired we may mark in quantities such as $a + 2\sigma$ $a + 3\sigma$, etc, etc

Absolute and Relative Measures of Dispersion

Two or more groups of objects may be compared by stating their respective means and dispersions Thus, for the two groups of estates given in Table 18 (p 72), we have—

State B $a = £13$ 750 $\sigma = £5$ 415

State C $a = £12,970$, $\sigma = £4$ 400

Frequency Distribution of Wages of Weekly Wage-earners showing various measures of Dispersion

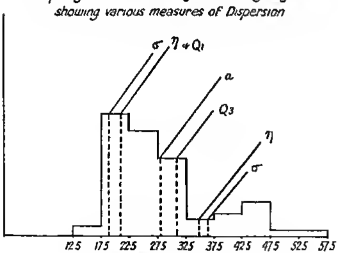


Fig 30

This method is satisfactory provided the two groups do not differ greatly in average size.

Should, however, the difference be considerable, it is better to determine the relative than the absolute amount of dispersion.

An industrial ordinary share priced at about £5 will vary more in an absolute sense than a share priced at about 5s., although it will probably vary less in a relative sense. In order to overcome this difficulty we may use a Coefficient of Dispersion found by dividing the measure by its appropriate central value. The coefficients accordingly are

$$\frac{\eta}{M}, \ \frac{\sigma}{a}, \text{ and } \frac{Q_3 - Q_1}{Q_3 + Q_1}$$[1]

The following table shows a selection of coefficients for the distribution of Estates given in Table 18.

TABLE 32

VALUES OF ESTATES—COEFFICIENTS OF DISPERSION

	η	$\frac{\eta}{M}$	σ	$\frac{\sigma}{a}$	$\frac{Q_3-Q_1}{2}$	$\frac{Q_3-Q_1}{Q_3+Q_1}$
State B . .	£ 4,625	0·31	£ 5,415	0·394	£ 4,610	0·35
State C . .	3,600	0·27	4,400	0·339	3,305	0·25

The quantity $\frac{100\sigma}{a}$ is known as the coefficient of variation and is denoted by CV.

[1] Note that σ, etc., are concrete measures, and should always be stated as so many units (shillings, years, etc.). $\frac{\sigma}{a}$ is an abstract quantity.

CHAPTER XII

SKEWNESS

The average and dispersion do not exhaust the information that the group is capable of supplying. There are also measures of skewness which indicate the tendency of the group to depart from symmetry. In ordinary language a skew distribution is lop-sided.

In a Symmetrical Distribution Mode, Median and Arithmetic Mean Coincide. Skewness has the effect of pulling the median and arithmetic mean away from the mode, sometimes to the right, sometimes to the left.

In Fig. 28, illustrating the distribution of wages of wage-earners, $Z = 22{\cdot}02s$, $M = 25{\cdot}66s$, and $a = 27{\cdot}85s$.

Provided the curve is not highly skewed the median usually travels *about* two-thirds the distance of the mean. Therefore approximately

$$M = Z + \tfrac{2}{3}(a - Z) \qquad (1)$$

In the above example

$$M = 22{\cdot}02 + \tfrac{2}{3}(27{\cdot}85 - 22{\cdot}02)$$
$$= 25{\cdot}91s$$

Actually $M = 25{\cdot}66s$, so that the agreement is fairly close in this case.

Graphical Illustrations

In Fig. 31 two distributions have been plotted with equal areas, means and standard deviations.

Distribution A (indicated by the firm line) is skewed and Distribution B (indicated by the dotted line) is symmetrical. The means of both curves coincide at the point $x = 0$. The mode of curve A falls at $x = -0{\cdot}4$ units. The standard deviation of both curves is $2{\cdot}236$ units.

Fig. 32 illustrates a distribution with a high degree of asymmetry. The curve is highest near the origin, falling rapidly and then more slowly. Technically it is known as a J shaped curve.

Coefficients of Skewness

Skewness is an abstract quantity and must be found in terms of the appropriate measure of dispersion.

The first coefficient is provided by the difference between arithmetic mean and mode, divided by the standard (or mean) deviation. This is the standard method.

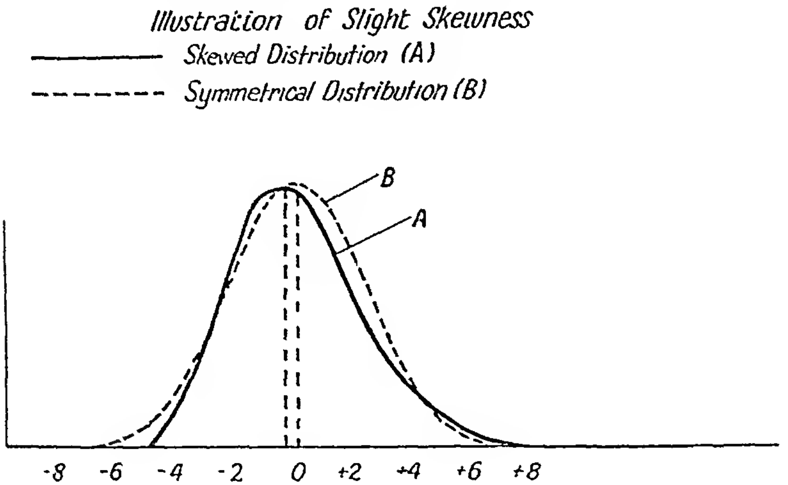


FIG. 31

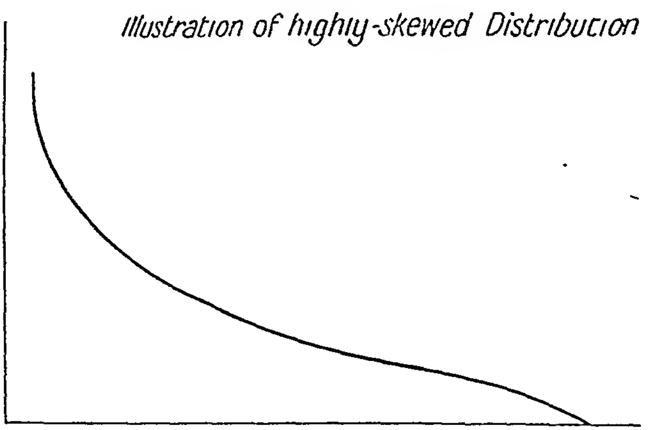


FIG. 32

Symbolically

$$j_1 = \frac{a - Z}{\sigma} \text{ or } \frac{a - Z}{\eta} \quad . \quad . \quad . \quad . \quad . \quad (2)^1$$

[1] This method is only satisfactory under certain conditions (with which the reader need not concern himself). If the mode is badly defined, we may use the alternative formula

$$= \frac{3\,(a - M)}{\sigma}.$$

If the arithmetic mean travels to the right there is positive and if to the left negative skewness symbolized by plus and minus signs respectively

In the wage distribution of Fig 19

$$j_1 = \frac{27.85 - 21.85}{8.85} = 0.68$$

In Fig 31 above

$$j_1 = \frac{0 - (-0.4)}{2.236} = 0.18$$

The second coefficient is based upon the fact that in a skewed distribution the median does not lie exactly half way between the Quartiles The appropriate coefficient is

$$j_2 = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} \tag{3}$$

In the wage distribution of Fig 19

$$j_2 = \frac{31.43 + 21.14 - 51.32}{31.43 - 21.14} = \frac{1.25}{10.29} = 0.12$$

Note the difference between the present figure and that yielded by the first method The two methods are based upon entirely different principles and their results are not comparable

The third coefficient is based upon the *third moment*[1] of the distribution

$$\text{The formula is } j_3 = \frac{\sqrt[3]{\dfrac{\Sigma f(x-a)^3}{n}}}{\sigma} \tag{4}$$

This measure is not used in elementary work

[1] A quantity of the form $m_r = \frac{\Sigma(fx^r)}{n}$ is known as the rth moment Usually moments are calculated around the arithmetic mean In that case we have

$$m_1 = \frac{\Sigma f(x-a)}{n} = 0$$

$$m_2 = \frac{\Sigma f(x-a)^2}{n} = \sigma^2 \text{ also known as the } \textit{variance}$$

$$m_3 = \frac{\Sigma f(x-a)^3}{n} \text{ etc. etc}$$

The theory of skewness is of more importance in biological and other investigations depending upon laboratory experiments than in the field of economic and social statistics.

The elementary student should familiarize himself with the definition of skewness and the various formulae employed. That is sufficient: he need not trouble to study the subject deeply.

CHAPTER XIII

PROBABILITY AND ERROR

I PROBABILITY

Statistical Method is ultimately based upon the Mathematical Theory of Probability The latter is one of the most difficult branches of mathematics and can be treated here only in elementary fashion

Definition of Probability

Probability is a measure of our expectation that an event will (or will not) happen

If an event can happen in m ways of which s represent successes and f failures, and each of these ways is equally likely the probability (or chance) of its happening is represented by

$$p = \frac{s}{m} \tag{1}$$

and of its not happening by

$$q = \frac{f}{m} \tag{2}$$

Evidently $s + f = m$ and $p + q = 1$ where 1 is taken as our measure of certainty

If a coin be tossed up the chance of a head is $\frac{1}{2}$ and of a tail $\frac{1}{2}$

The probability of drawing an ace from a pack of 52 cards is $\frac{4}{52} = \frac{1}{13}$

Compound Probabilities

The chance of concurrence of two or more independent events is given by the product of their chances of separate occurrence In the coin tossing experiment the chance of obtaining 3 heads in 3 throws is

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = (\tfrac{1}{2})^3 = \tfrac{1}{8}$$

This case must be contrasted with the case in which the chances are not independent. The chance of drawing 4 aces from a pack of 52 cards when the cards are not replaced is $\frac{4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49}$, since each draw affects the chance of the next.

If two coins are tossed up instead of one, there are altogether 4 ways[1] in which they may fall, viz.—

HH,	with a chance of	$\frac{1}{4}$
HT, TH,	,, ,,	$\frac{1}{2}$
TT,	,, ,,	$\frac{1}{4}$
		1

With three coins there are 8 ways,[2] viz.—

HHH,	with a chance of	$\frac{1}{8}$
HHT, THH, HTH,	,, ,,	$\frac{3}{8}$
HTT, THT, TTH,	,, ,,	$\frac{3}{8}$
TTT,	,, ,,	$\frac{1}{8}$
		1

The above series evidently form frequency distributions with a law following the binomial theorem. It is easy to prove that if n coins be tossed up, the frequency distribution of the respective chances of $n, n-1, n-2 \ldots 2, 1, 0$ heads is given by the expansion of $(\frac{1}{2} + \frac{1}{2})^n$.

The experiment need not be confined to coin tossing, nor need the chances of success and failure be exactly equal. If p and q be any proper fractions representing the chances of success and failure for a single event ($p + q = 1$), the frequency distribution of the chances of $n, n-1, n-2 \ldots 2, 1, 0$ successes in the compound event is given by the successive terms of $(p+q)^n$, expanded by the binomial theorem.[3]

The following Table exhibits the results of expanding the expression $10{,}000\ (q + p)^{20}$ for $p = 0{\cdot}1$, $0{\cdot}3$, and $0{\cdot}5$ respectively.

[1] There are 2 events, each of which may happen in 2 ways. Total $2^2 = 4$ ways

[2] There are 3 events, each of which may happen in 2 ways. Total $2^3 = 8$ ways

[3] See any textbook on Elementary Algebra.

TABLE 33

TERMS OF THE BINOMIAL SERIES $10\,000\,(q + p)^{20}$ [1]

(Figures to nearest unit)

No of Successes	Frequency		
	$p = 0\cdot1$ $q = 0\cdot9$	$p = 0\cdot3$ $q = 0\cdot7$	$p = 0\cdot5$ $q = 0\cdot5$
(1)	(2)	(3)	(4)
0	1 216	8	
1	2 702	68	
2	2 852	278	2
3	1 901	716	11
4	898	1 304	46
5	319	1 789	148
6	89	1 916	370
7	20	1 643	739
8	4	1 144	1 201
9	1	654	1 602
10		308	1 76
11		120	1 602
12		39	1 201
13		10	739
14		2	370
15			148
16			46
17			11
18			2
19			
20			
$a = pn$	2	6	10
$\sigma = \sqrt{pqn}$	1 3416	2 0494	2 2361

Normal Frequency Curve

If n (the number of events) be large and neither p nor q very small the expansion of $(q + p)^n$ approximates to a regular curve known as the Normal Frequency Curve [2] This curve represents the standard frequency distribution of a variable which depends upon the combination of independent chances A drawing of the curve is given below It is symmetrical rising to a high peak in the centre and tailing off to infinity in both directions

[1] *Cf* Yule and Kendall *Theory of Statistics* (1937) p 181

[2] The formula of the curve is $y = Ce^{-\frac{x^2}{2\sigma^2}}$ where x represents the distance of the variable from the centre of the distribution σ the standard deviation and C a constant.

The origin (or zero point) is taken in the centre, the variable is measured in a positive or negative direction along the x axis, and its frequency along the y axis. The chance of a deviation as great as d is evidently proportional to the area included between the curve,

Normal Frequency Curve. (Normal Curve of Error)

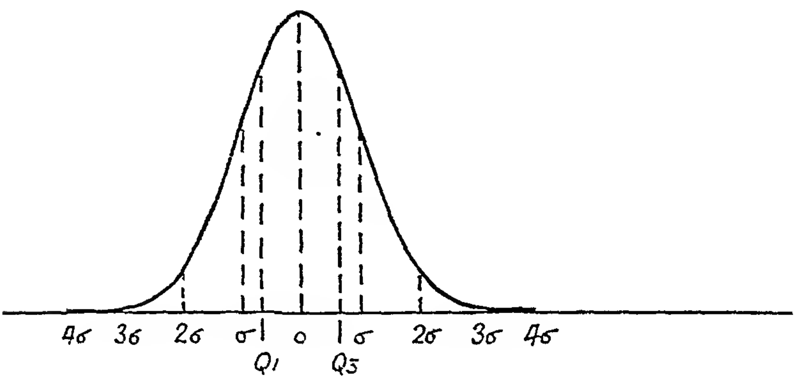


Fig. 33

the axis of x and the ordinates through the points $x = d$ and $x = -d$. Tables of the curve have been prepared which considerably lighten the work of calculation. A summary table is given below—

TABLE 34

CHANCES OF DEVIATION FROM THE CENTRE ACCORDING TO THE NORMAL FREQUENCY CURVE

Deviation Lying Between (1)			Corresponding Chance (2)
$+ 0{\cdot}5\sigma$	and	$- 0{\cdot}5\sigma$	0·383
$+ 0{\cdot}6745\sigma$	„	$- 0{\cdot}6745\sigma$	0·500
$+ 1{\cdot}0\sigma$	„	$- 1{\cdot}0\sigma$	0·682
$+ 1{\cdot}5\sigma$	„	$- 1{\cdot}5\sigma$	0·866
$+ 2{\cdot}0\sigma$	„	$- 2{\cdot}0\sigma$	0·954
$+ 2{\cdot}5\sigma$	„	$- 2{\cdot}5\sigma$	0·988
$+ 3{\cdot}0\sigma$	„	$- 3{\cdot}0\sigma$	0·997

Chances as Indicated by the Normal Curve.

From this table it appears that for a variable (x) conforming to the Normal Law, there is a chance of 682 in 1,000 that a deviation

will not exceed the standard deviation of the curve (σ) and a chance of 318 in 1 000 that it will There is a chance of 997 in 1 000 that a deviation will not exceed 3 times the standard deviation (3σ) and a chance of only 3 in 1 000 that it will

Since the curve is symmetrical arithmetic mean median and mode coincide The two **Quartiles** are located at the points

$$x = 0\ 6745\sigma \text{ and } x = -0\ 6745\sigma$$

Since the Quartiles enclose half the area of the distribution the chance of a deviation (d) not exceeding this amount is $\frac{1}{2}$ The quantity 0 6745σ is known as the **Probable Error**

Normal Distributions in Practice

The theory of the Normal Curve is not confined to experiments of the type described Developments under more complicated conditions are possible which may be crystallized in the statement that where a variable represents the sum (or average) of a large number of elementary variables its distribution will tend towards normality provided the elementary variables are independent or nearly so

In practice the above conditions are seldom realized in their entirety and straightforward normal distributions of physical objects are rare except in certain branches of biological science The Normal Frequency Curve is interesting chiefly in connection with the theory of compensating errors which may be regarded as independent chance events with a tendency to compound themselves upon the principles just described In this case the curve is known as the **Normal Curve of Error**

A Priori and Empirical Probabilities

The chances so far described have been of the *a priori* kind i e chances deduced from consideration of the possible combinations of events involved In the majority of cases however the factors at work are not definitely known and the conditions are too complicated to enable us to isolate the elementary chances involved We must therefore fall back upon empirical probabilities, based upon actual observation and experiment

Let m trials be made of which s represent successes and f

failures. Upon this basis our best estimate of the chance of the event happening is

$$p' = \frac{s'}{m'} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

and of its not happening is

$$q' = \frac{f'}{m'} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3b)$$

The larger m' (the number of trials), the more accurate will be our estimate. Since, however, m' is necessarily finite, the whole experiment is only a sample, and were it repeated several times, slightly different values of p' and q' would result. By increasing m', the error in the determination of p' and q' may be made as small as we please.

According to the *Life Table*, out of every 100,000 persons living at age 10, 82,277 survive to age 40, of whom 823 die during the year and 81,454 survive until age 41.

Therefore, the chance of a life aged 40 dying within the year is

$$q_{40} = \frac{823}{82{,}277} = 0{\cdot}01000$$

and the chance of his surviving the year is

$$p_{40} = \frac{81{,}454}{82{,}277} = 0{\cdot}99000$$

Of these 81,454 persons, 846 die within the next year and 80,608 survive until age 42.

Therefore $$q_{41} = \frac{846}{81{,}454} = 0{\cdot}01038$$

and $$p_{41} = \frac{80{,}608}{81{,}454} \doteq 0{\cdot}98962$$

The chance of a life aged 40 surviving until 42 is evidently represented by the compound chance of a life aged 40 surviving until 41 and of a life aged 41 surviving until 42, i.e.

$$p_{40} \times p_{41} = \frac{80.608}{82{,}277} = 0{\cdot}97971$$

and the chance of his dying $= 1 - 0{\cdot}97971 = 0{\cdot}02029$

These calculations can evidently be extended to cover any combination of probabilities desired

The Rule that empirical probabilities may be estimated from statistical frequencies must be applied with considerable discretion The notion of Probability implies underlying stability and fixation, subject only to the fluctuations engendered by passing disturbances There must be some fundamental cause (or complex of causes) persistent in time and space and operating in accordance with a regular law The data must be numerous and homogeneous and the general conditions of the problem must remain the same during the whole series of experiments

There is evidently more likelihood of securing these conditions as regards natural phenomena than as regards economic and social phenomena which are subject to arbitrary interference at man's hands

II Statistical Errors

A Statistical Error is a measure of the difference between our estimate of a quantity and its true value [1]

Statistical measurement is seldom precise As a rule its data are not susceptible of hard and fast definitions nor are they under control Unlike the chemist or physicist the statistician cannot experiment he can only observe and record the results Moreover statistical method does not aim at mathematical precision but at broad views in which confusing and unnecessary detail is rigorously suppressed Human judgment is not susceptible to small differences in objects and time spent in calculating to a higher degree of accuracy than is actually warranted is time wasted

The Statistician thinks in terms of probabilities and estimates and is satisfied with reasonably accurate results provided he can also measure their reliability

Sources of Statistical Errors

The following are the chief sources of statistical errors—

1 Errors of Origin—Bias in information collected lack of definition in the subject matter of inquiry Erratic tendencies inherent in the data themselves

2 Errors of Inadequacy—Lack of representativeness due to smallness of sample studied

[1] It is not a mistake but a difference arising from any source of inexactitude

3. Errors of Manipulation—Involuntary mistakes in counting, measurement, description, classification, and approximation; arithmetical and clerical errors.

Measurement of Error.

The Absolute Error of an estimate is the difference between the estimate and the true value, as above described.

Let x represent the estimate

x' „ the true value

e „ the error

Then $e = x - x'$ . . . . (4a)

And $x = x' + e$ . . . . (4b)

The Relative Error is the ratio of the Absolute Error to the estimate (or to the true value).

$$\varepsilon = \frac{e}{x} \text{ or } \frac{e}{x'} = \text{relative error} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Theoretically the true value should be chosen as divisor, but if this be unknown, an estimate may be substituted. The difference between the two methods may usually be neglected.

The Absolute Error is significant when it is a question of addition or subtraction, and the Relative Error when it is a question of multiplication or division.

Possible, Standard, and Probable Errors.

The Possible Error represents the maximum and minimum limits between which the actual error must lie.

If a quantity be rounded off to the nearest 100, the upper limit of error is $+50$ and the lower limit -50. The Possible Error is therefore represented by ± 50.

In this particular problem the law of distribution of error is known exactly, and the limits may be definitely ascertained by theory. As a rule, however, the law is only known approximately, or not at all, and there are no definite limits. In such cases we must be content with a rough estimate of the possible error, defining it as the pair of limits outside which errors are extremely unlikely to occur. Owing to this lack of definition, it is preferable

to utilize the Standard Error, which may be estimated with sufficient precision when the number of items is large enough

The Standard Error of an estimate is equivalent to the error s standard deviation

This measure implies that the quantity s law of error is known i e that the respective chances of errors of various sizes can be expressed as a frequency distribution

In the rounding-off problem given above the actual error may assume any value between $+50$ and -50 and each of these values i equally likely The frequency d stribution of the chances of error is therefore given by a rectangle

It is easy to prove that the standard deviation of this rectangle is

$$\sqrt{\frac{100^2}{12}} = 28\ 87$$

The Probable Error of an estimate is equivalent to the error s Quartile Deviation

In the rounding off problem the two Quartiles are located at tl e points $+25$ and 25 and the Probable Error is ± 25 In other words the chance of the error lying between the limits ± 25 is $\frac{1}{2}$

The use of the Probable Error is going out of fashion Modern statisticians use the Standard Error instead

Biased and Unbiased Errors

Biased (or Cumulative) Errors If a quantity is such that its errors all tend to l e in the same direction the error is said to be biased (or cumulative)

Stock Exchange securities are biased upwards during a boom and downwards during a slump In the former case the market tends to over-estimate their value as expressed by the earning capacity of the companies concerned and in the latter case to under estimate

Unbiased (or Compensating Errors) If a quantity is such that its errors tend to neutralize one another the error is said to be unbiased (or compensating)

In normal times some Stock Exchange securities are overvalued and some undervalued on their prospects In the mass therefore they tend to be fairly valued

Laws of Error.

As a rule the actual error committed with respect to any given estimate is and must remain unknown. It is only in special cases that the true value can be determined and the actual error checked up. This fact need not, however, disturb us, provided we know something of the properties of the class of error in question.

As a rule, however, it is more important to know the typical value of the error than its value upon any particular occasion. This typical value can be estimated if the error in question follows a known law.

Errors Following the Normal Law.

Errors of Observation (when Unbiased) and other Errors depending upon Chance Combinations tend to follow the Normal Law. Their probable frequency distribution can therefore be found, provided the standard error can be measured or estimated. The following table supplies an illustration—

TABLE 35

FREQUENCY DISTRIBUTION OF 1,000 OBSERVATIONS OF THE RIGHT ASCENSION OF *Polaris*[1]

Deviation of Observation from – 0·06 seconds (1)	Actual Frequency (2)	Theoretical Frequency (3)
Seconds		
– 3·5	2	4
– 3·0	12	10
– 2·5	25	22
– 2·0	43	46
– 1·5	74	82
– 1·0	126	121
– 0·5	150	152
0	168	163
0·5	148	147
1·0	129	112
1·5	78	72
2·0	33	40
2·5	10	19
3·0	2	10
	1,000	1,000

A drawing of the distribution is given on p. 124.

[1] Whittaker and Robinson, *Calculus of Observations*.

The theoretical figures as determined by the Normal Curve of Error are indicated by the dotted line. The correspondence between the two figures is remarkably close.

In practice we do not usually go to the trouble of calculating the

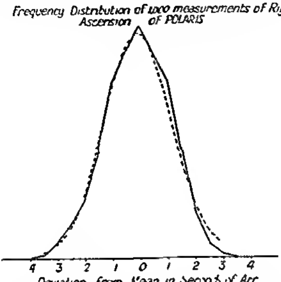


FIG. 34

theoretical curve but simply test whether the distribution as a whole falls within the spread marked out by expectation.

The constants of the theoretical curve are—

$$a = -0{\cdot}06 \text{ sec} \quad \text{and}$$

$$\sigma = 1{\cdot}1785 \text{ sec},$$

and upon the basis of Table 34 we should expect 99·7 per cent of the observations to lie between the limits $-0{\cdot}06 \pm 3 \times 1{\cdot}1785$, i.e. $-3{\cdot}60''$ and $3{\cdot}48''$. In fact the whole distribution lies between these limits, so that the spread is slightly less than that given by theory.

Regarding a chance of 0·3 per cent as negligible we deduce the traditional rule that any difference between the expected value and

its corresponding actual value exceeding (say) three times[1] the standard error (or four times the probable error) is statistically significant. In other words, it is too large to be accounted for by chance, and the presence of some specific cause must be inferred.

Chances of Error Equal.

In the case of errors of rounding-off and in some other cases, the chances of error over a given range are equal. The frequency distribution is therefore given by a rectangle and the probable error by a magnitude equal to one-quarter of the range.

In this form of distribution the chances of error do not decrease as the error increases. Consequently, the range of the probable error is widened, so that it equals one-half of the possible error.

Other Distributions.

The Laws of Error indicated above are the most important. Other Laws can be developed, but the discussion would take us outside the limits of this book. If the error under consideration cannot be brought under either of the Rules given above, it is the best plan to work from the Possible Error, which should be estimated on the best information available.[2]

Propagation of Errors.

In the following analysis the x's represent quantities subject to error, and the e's their possible errors. Both x's and e's are construed in their *algebraic* sense.[3]

Errors combine according to the ordinary Laws of Algebra, i.e.

$$(x_1 + e_1) + (x_2 + e_2) = (x_1 + x_2) + (e_1 + e_2) \quad . \quad . \quad . \quad (6)$$

$$(x_1 + e_1) - (x_2 + e_2) = (x_1 - x_2) + (e_1 - e_2) \quad . \quad . \quad . \quad (7)$$

$$(x_1 + e_1)(x_2 + e_2) = x_1 x_2 (1 + \varepsilon_1)(1 + \varepsilon_2) =$$

$x_1 x_2 \{1 + (\varepsilon_1 + \varepsilon_2)\}$ approximately, if the product $\varepsilon_1 \varepsilon_2$ is so small that it may be neglected . . . . . (8)

[1] Some statisticians take twice the standard error as their criterion (instead of three times).

[2] It is usually sufficient if one can make a good guess at the order of magnitude in question. This guess may be based upon the nature of the data or upon past experience in dealing with similar data.

[3] A quantity may be algebraically positive whilst arithmetically negative.

$$(x_1 + e_1) \div (x_2 + e_2) = \frac{x_1}{x_2}(1 + \varepsilon_1)(1 + \varepsilon_2)^{-1}$$
$$= \frac{x_1}{x_2}\{1 + (\varepsilon_1 - \varepsilon_2)\} \text{ approximately neglecting squares and products} \tag{9}$$

Examples

$$(200 + 5) + (100 + 2) = 300 + 7 \tag{6a}$$
$$(200 + 5) - (100 + 2) = 100 + 3 \tag{7a}$$
$$(200 + 5)(100 + 2) = 200 \times 100\,(1 + 0{\cdot}025)(1 + 0{\cdot}020)$$
$$= 20000\,(1 + 0{\cdot}045) = 20\,000 + 900 \text{ (approximately)} \tag{8a}$$
$$(200 + 5) \div (100 + 2) = \frac{200}{100}(1 + 0{\cdot}025)(1 + 0{\cdot}020)^{-1}$$
$$= 2(1 + 0{\cdot}005) = 2 + 0{\cdot}010 \text{ (approximately)} \tag{9a}$$

Equations (6) to (9) must be construed algebraically In examples *(6a) to (9a) it has been assumed that the es are definitely cumulative* and have the same signs as the xs This is not however a necessary condition Assume for example that e_2 is essentially negative

On this basis we have

$$(200 + 5) + (100 - 2) = 300 + 3 \tag{6b}$$
$$(200 + 5) - (100 - 2) = 100 + 7 \tag{7b}$$
$$(200 + 5)(100 - 2) = 20\,000\,(1 + 0{\cdot}005)$$
$$= 20\,000 + 100 \text{ (approximately)} \tag{8b}$$
$$(200 + 5) \div (100 - 2) = 2(1 + 0{\cdot}045) = 2 + 0{\cdot}090 \text{ (approximately)} \tag{9b}$$

Now assume that the *sizes* of the es can be estimated but that their *signs* are not definitely known In this case we must seek the most unfavourable combinations of errors that can occur and we accordingly rewrite equations (6) to (9) as follows—

$$(x_1 \pm e_1) + (x_2 \pm e_2) = (x_1 + x_2) \pm (e_1 + e_2) \tag{6c}$$
$$(x_1 \pm e_1) - (x_2 \pm e_2) = (x_1 - x_2) \pm (e_1 + e_2) \tag{7c}$$
$$(x_1 \pm e_1)(x_2 \pm e_2) = x_1 x_2\{1 \pm (\varepsilon_1 + \varepsilon_2)\} \tag{8c}$$
$$(x_1 \pm e_1) \div (x_2 \pm e_2) = \frac{x_1}{x_2}\{1 \pm (\varepsilon_1 + \varepsilon_2)\} \tag{9c}$$

EXAMPLES

$$(200 \pm 5) + (100 \pm 2) = 300 \pm 7 \quad . \quad . \quad . \quad . \quad . \quad (6d)$$

$$(200 \pm 5) - (100 \pm 2) = 100 \pm 7 \quad . \quad . \quad . \quad . \quad . \quad (7d)$$

$$(200 \pm 5)\,(100 \pm 2) = 20{,}000\,(1 \pm 0{\cdot}045) = 20{,}000 \pm 900 \quad . \quad (8d)$$

$$(200 \pm 5) \div (100 \pm 2) = 2(1 \pm 0{\cdot}045) = 2 \pm 0{\cdot}090 \quad . \quad . \quad (9d)$$

Numbers Large and Errors Compensating.

If the number of quantities be large and their errors of the compensating variety, there will be a reduction of error due to mutual cancellations. In this case our best estimates for the Errors of Sums or Differences are of the form

$$\pm \sqrt{e_1^2 + e_2^2 + \ldots + e_n^2}$$

or (say) $\pm \bar{e}\sqrt{n}$, where $\bar{e}$ is typical of the e's . . . . (10)

This rule applies both to Standard and to Possible Errors.

Results of Discussion.

The results of this discussion show that the absolute error of a sum (or difference) involves the sum (or possibly the difference) of the absolute errors of its components, and that the relative error of a product (or quotient) involves the sum (or possibly the difference) of the relative errors of its components.

Absolute biased error may be reduced by subtraction and relative biased error by division of the quantities.

If the number of quantities is large and their errors compensating, the resultant error of their sum (or difference) is *reduced in the ratio* represented by the fraction $\frac{1.}{\sqrt{n}}$

CHAPTER XIV

SAMPLING

THE Theory of Sampling deals with the conditions under which statistical induction is possible and supplies the necessary tests Since statistical science is quantitative such tests are in general more precise than those available in other branches of knowledge

No one has time or means to make a complete investigation of every problem with which he comes into contact he must therefore proceed by sample

In the ordinary affairs of life man acts upon this principle un thinkingly Science acts upon the same principle only more cau tiously Organized knowledge is representative in character In business agreements it is often provided that the goods supplied shall be in accordance with sample Accountants make test checks of their clients books Social investigators make sample inquiries into the conditions of representative households

General Laws of Statistical Induction

The two General Laws governing processes of statistical induction are—

1 **The Law of Statistical Regularity** which lays down that a group of objects chosen at random from a larger group tends to possess the characteristics of that larger group [1]

The sample group must not be unreasonably small for it would then be excessively affected by abnormal items The sample must be chosen at random Every item in the population must stand an equal chance of being included The only safe method of securing perfect randomness is to draw lots

The Larger the Sample the More Reliable are its indications It will be shown later that its reliability is proportional to the square root of the number of items included.

In general the results of sample inquiries will show differences that cannot be assigned to any definite cause Every sample will have its own peculiarities reflected in the form of its frequency distribution and in the precise magnitudes of its average standard

[1] *The larger group is known as the* ***Population*** (See p 75)

deviation, and skewness. These differences are known as fluctuations, and it is the aim of the Theory of Sampling to supply tests whereby to determine whether any given fluctuation is statistically significant or not.

Two or more samples may be merged and the result treated as a single sample. If a number of samples be collected, they may be regarded as composite observations, and their characteristics described by frequency distributions, averages, etc.

2. The Law of Inertia of Large Numbers asserts that large aggregates are relatively more stable than small ones. The movements of an aggregate are the resultants of the movements of its separate parts; and it is improbable that the latter will all be moving in the same direction at the same time. Consequently their movements will tend to compensate one another, and the larger the numbers involved, the more complete will this compensation be. This law provides the large-scale counterpart of the Law of Compensating Errors.[1]

Both the Law of Statistical Regularity and the Law of Inertia of Large Numbers are based upon experience. They may be exemplified by insurance statistics. The actuaries collect large masses of data from past experience, on the basis of which they calculate the chances of occurrence of various contingencies forming the subjects of insurance, and thence the premiums to be charged. Although the company's experience will cover a large field, extending back many years, it is only a *sample* experience, minute in comparison with the totality of events not recorded. Nevertheless, it will reproduce the characteristics of the whole, and can be used with confidence for the fixation of the company's tariffs.

The amount of the claims that will be made by any one individual or small group of individuals will vary considerably from year to year and cannot be predicted, whilst the total for a relatively large group can be predicted with some success, and the total business of the company over a period of years will show a high degree of stability.

In the same way the total of the world's production of crops fluctuates less than the figures for individual areas, for bad harvests in one part of the world are usually compensated to some extent by good harvests in other parts.

[1] See Chapter XIII, p. 122.

Secular Changes

The Law of Inertia of Large Numbers does not of course preclude secular movements due to long period tendencies in the background of conditions The birth rate of England and Wales is relatively stable from year to year but for a long period it has shown a progressive decline the causes of which are a matter of controversy

Errors of Sampling

Errors of origin and errors of manipulation are common to all kinds of statistical work Inquiry by sample involves a special kind of error due to the fact that no sample affords a perfect representation of the population from which it is drawn In spite of every precaution to secure randomness slight variations must emerge due to the elements of chance present in every selection Such variations are known as fluctuations of sampling and upon their probable magnitude depends the reliability of the sample

Attention has already been drawn to the distinction between Quantitative and Qualitative Variates If the characteristic studied is non measurable the sample inquiry aims at determining the proportion of the objects in the population possessing that characteristic If the characteristic is measurable the sample inquiry aims at determining its mean dispersion and skewness

Qualitative Variate—Standard Error of a Proportion

In a sample inquiry it is found that s items out of n possess a given characteristic (C) The observed proportion is $p = \frac{s}{n}$ What is the expectation that p will apply to the total population?

The best plan is to start with the population (supposed known)

Let there be a homogeneous population of N objects of which pN possess a given characteristic and qN do not ($p + q = 1$)

Draw a sample of n[1] items at random Reckoning the presence of C as a *success* and its absence as a *failure* the respective chances of 0 1 2 $r - 2$ n 1 n *successes* in the sample will be given by the successive terms of the expansion of $(q + p)^n$ These terms may be represented by a frequency distribution with mean pn

[1] n is supposed to be small with respect to N

and standard deviation $\sqrt{pqn}$.[1] Upon this basis the mathematical expectation of the number of successes in a random sample of n is measured by

$$pn \pm \sqrt{pqn}$$

where the second term represents the standard error or measure of unreliability of the first . . . . (1)

Division by n gives the expectation of the proportion of successes as

$$p \pm \sqrt{\frac{pq}{n}} \quad . \quad . \quad . \quad . \quad (2)$$

In other words, the chance of the observed proportion (p') differing from p by a quantity as great as $\sqrt{\frac{pq}{n}}$ is about 0·3, and the chance of the deviation being three times as large is negligible.[2]

Suppose, now, we have no information as to the magnitude of p (the proportion in the population) beyond what is afforded by the sample.

In these circumstances our best estimate of p is given by the equation.

$$p = p' \pm \sqrt{\frac{p'q'}{n}} \quad . \quad . \quad . \quad (3)$$

where the second term on the right affords a measure of the unreliability of our estimate.

The reliability increases with the square root of n.

The total numbers of legitimate births registered in the *Midland* registration areas during 1929 were: male, 34,897; female, 33,609; total, 68,506. What is the estimated proportion for the total population?

Here

$$p' = \frac{34{,}897}{68{,}506} = 0{\cdot}5094$$

$$q' = \frac{33{,}609}{68{,}506} = 0{\cdot}4906$$

$$\sqrt{\frac{p'q'}{n}} = 0{\cdot}0019 = \sigma_p'$$

[1] The student may test this statement by working out the mean and standard deviations of the distributions given in Table 30.

[2] See Chapter XIII, p. 125.

Upon this basis the estimated proportions of births in the population are—

Male, 0.5094 ± 0.0019
Female, 0.4906 ± 0.0019

We may check up this result by reference to the total numbers of legitimate births registered in England and Wales in 1929, viz —

Male, 313 836, *female*, 300,530, total, 614 366

Here
$$p = \frac{313\,836}{614\,366} = 0.5108$$

$$q = \frac{300\,530}{614\,366} = 0.4892$$

$p - p' = 0.0014 = 0.74\sigma$ The difference is less than the standard error of p' [1]

Therefore the difference between p and p' is not statistically significant, and the estimate of p' from the sample population is reliable

Quantitative Variate—Standard Error of Arithmetic Mean

In a sample inquiry the mean value of a certain variable (x) is found to be a What is the expectation that a will apply to the whole population?

This problem is more complicated than the previous one It is no longer a question whether the characteristic is present, but how much it is present

Let there be a homogeneous population of N objects varying in size Draw n items at random and record the values of the 1st, 2nd nth items drawn together with their aggregate value as follows—

$$x_1 + x_2 \quad + x_n = X = na$$

where x_1 represents the size of the first item drawn, x_2 the size of the second, etc etc , X represents the sum of the x's and a their mean size

Repeat this process m times in all

Owing to the randomness of selection the values of x_1 x_2 x_n will vary considerably from sample to sample Their sum (X) and

[1] Actually the chance of a deviation exceeding $0{\cdot}74\sigma$ is 46 0 per cent

their mean (a) will also vary, but to a less marked extent because the variations of the x's will tend to compensate one another.

The successive sample values of a can therefore be plotted as a frequency distribution, provided the dispersion of a round its mean can be found.

It has been shown that the variance of the sum of n quantities equals the sum of the variances of its components, provided the latter are free to vary independently.[1]

Upon this basis

$$\sigma_x^2 = \sigma_1^2 + \sigma_2^2 \ldots + \sigma_n^2 = n^2\sigma_a^2$$

where σ_x represents the standard deviation of the sum of the variables; $\sigma_1, \sigma_2 \ldots \sigma_n$, the standard deviations of the 1st, 2nd ... nth items drawn in the samples, and σ_a the standard deviation of the average.

Since the x's are drawn at random their dispersions will tend towards equality with a value (σ) equivalent to the standard deviation of x in the whole population.

Therefore we may write

$$\sigma_x^2 = n\sigma^2 = n^2\sigma_a^2$$

Whence

$$\sigma_a = \frac{\sigma}{\sqrt{n}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where σ_a represents the standard error of the mean, and σ the standard deviation of the individual items.

Therefore, our best estimate of the mean of the population is

$$a \pm \sigma_a = a \pm \frac{\sigma}{\sqrt{n}} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Upon this basis the standard error of the mean wage in Table 25 is given by

$$\frac{8{\cdot}85}{\sqrt{72}} = 1{\cdot}05\text{s.}$$

and the mean of the distribution should be written

$$27{\cdot}85 \pm 1{\cdot}05\text{s.}$$

Here the second quantity is a measure of the unreliability of the estimation of the mean wage from a sample of seventy-two wage earners only.

[1] See Chapter XI, p. 103.

Distribution of Arithmetic Mean

The means of random samples of n items from a normal population are normally distributed with standard error

$$\frac{\sigma}{\sqrt{n}} \qquad (6)$$

If the parent population is non normal the distribution of means will still tend to normality if n be sufficiently large In other words if instead of plotting single observations we plot means of random samples of n we should get a normal curve with standard deviation reduced in the ratio in question

Small Samples

As a rule the true population standard deviation is unknown and as a makeshift we have to use an estimate based on values given by samples When the number of items in the samples is large the error is inappreciable but with small samples a bias is introduced into the calculations to correct which special methods are necessary [1]

Other Statistical Quantities

The standard errors of other statistical quantities can be calculated but the processes are tedious and complex.

[1] Consult Fisher's *Statistical Methods for Research Workers* or Tippett's *Methods of Statistics*

CHAPTER XV

CORRELATION

The study of quantitative relationships reaches its fullest expression in sciences like astronomy, chemistry, mechanics, and physics, whose data are susceptible of exact measurement and whose problems are capable of formulation upon a strict mathematical basis. The application of quantitative methods to sciences like biology and sociology has necessitated the development of a special technique, in which elasticity is gained at the expense of precision. The most important developments in the study of statistical relationships are embodied in the theory of correlation.

Definition of Correlation.[1]

If two or more quantities vary in sympathy, so that movements in the one *tend* to be accompanied by corresponding movements in the other(s), then they are said to be correlated.

It is implied that the relationship cannot be expressed in definite mathematical form, for if it could, then statistical methods would be unnecessary.

Problems Involved.

The problems involved by Correlation are two in number—

1. The measurement of the amount of dependence between the variables that can be accounted for.

2. The expression of the probable movements of one variate (the relative) in terms of the other (the subject).

Correlation may be direct or inverse; direct when the variables tend to move in the same sense, and inverse when they tend to move in opposite senses.

Correlation Admits of Degrees. We may detect not only its presence, but the amount of its presence. Perfect correlation implies a rigid connection between the variables, typified by a mathematical function, whilst imperfect correlation implies a looser connection which finds expression in terms of probability. Denoting

[1] Sometimes referred to as *Co-variation*.

perfect correlation by $+1$ or -1 according as it is direct or inverse complete independence is denoted by o and limited dependence by intermediate values of the coefficient

Correlation and Mathematical Functionality

In mathematical analysis a function is typified by $y = f(x)$ where x represents the independent and y the dependent variable From our present standpoint the essential feature of such a function is that y is completely determined when x is given and *vice versa* There is nothing left over

The statistical counterpart of this expression may be written $y = \phi(x) + v$ where $\phi(x)$ represents the most probable[1] value of y for a given x and v a residual or difference between y as computed and y as observed The residual (v) implies the presence of disturbing factors that can neither be analysed nor accounted for It is analogous to a statistical error and obeys the same laws

Multiple Correlation

Correlational methods are not limited to two variables Theoretically it is possible to introduce any number of variables and to suppose them connected in any ways that may appear appropriate Such developments of the subject are however too complicated for an elementary textbook and we shall confine ourselves to the problem of two variables connected by linear relationships This will afford the student a useful introduction to the subject and enable him to pursue it in more specialized treatises

A Simple Illustration

The problem of correlation between the ages of husband and wife affords a useful illustration of the problems involved The dominant biological and sociological factors are obvious and need not be elaborated These are however liable to disturbance by other factors too numerous and too complicated to classify with the result that the relation between the two variables is only approximative

Consider the following data—

[1] Or alternatively the average value or best estimate

TABLE 36

CORRELATION BETWEEN AGES OF HUSBANDS AND WIVES AT MARRIAGE[1]

Husband's Age (X) (1)	Wife's Age (Y) (2)
Years	Years
23	18
27	20
28	22
28	27
29	21
30	29
31	27
33	29
35	28
36	29

Scatter Diagram.

The relationship may be exhibited in the form of a scatter diagram (Fig. 35) in which the ten representative points are located by reference to the rectangular axes X (husband's age) and Y (wife's age).[2] The observations lie in the form of a belt, sloping upwards and to the right. Hence we infer the instance of correlation.

Contrast this figure with Fig. 36, representing pairs of numbers taken at random from mathematical tables. The two variables are scattered fairly uniformly over the four quadrants, and there is no tendency for them to vary in sympathy.

Regression Lines.

When the number of observations is large, the diagram takes the form of a swarm of points and the existence of correlation may be inferred from the local density of the swarm. Owing, however, to the excessive labour involved by plotting the points separately, it is better to draw up a correlation table (Table 40) or a Regression diagram (Fig. 38 on page 148).

A line showing the average values of one variate associated with

[1] Society of Incorporated Accountants, November, 1930.

[2] In this Chapter the symbol—

X represents the observed value of the variable;

x represents the deviation from its arithmetic average;

x' represents the deviation from any other point, e.g. a trial average.

Correlation between Ages of Husband & Wife
Scatter Diagram

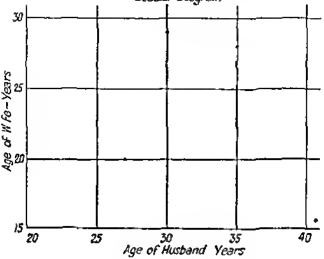


FIG 35

Case of Independence

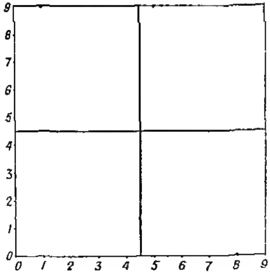


FIG 36

given values of the other is known as a Regression Line. Fig. 38 shows the regressions of the data in Table 40. There are two regression lines, one showing the average value of Y for a given X, and the other the average value of X for a given Y. The two *loci* diverge somewhat. This is a consequence of the correlation not being perfect.

Provided the number of observations is sufficiently large, the Regression Lines will usually follow definite curves, with a tendency to become erratic at one or both ends, where the number of observations is small. Sometimes the Regression is approximately linear, and in that case we may be justified in substituting a straight line for the curve running through the successive means of arrays. The *locus* of this line may be found approximately by stretching a thread across the diagram. The operation is, however, uncertain and, where accuracy is desired, we must resort to mathematical methods.

The Correlation Coefficient.

Transfer the origins of X and Y to their respective means, i.e. consider the deviations $x = X - \bar{X}$ and $y = Y - \bar{Y}$ instead of the original observations.

Now consider the quantity $p = \dfrac{\Sigma(xy)}{n}$, known as the mean product of X and Y.

Some of the x's are plus and others minus, and the same remark applies to the y's. If X and Y are independent, pairs of plus values and minus values will occur indiscriminately, and p will tend towards zero; but if X and Y are *not* independent, there will be a prevailing tendency for

$$\text{positive } x\text{'s to be associated with} \begin{Bmatrix} \text{positive } y\text{'s} \\ \text{negative } y\text{'s} \end{Bmatrix}$$

$$\text{and negative } x\text{'s to be associated with} \begin{Bmatrix} \text{negative } y\text{'s} \\ \text{positive } y\text{'s} \end{Bmatrix}$$

$$\text{according as } X \text{ and } Y \text{ tend to move in} \begin{Bmatrix} \text{the same sense} \\ \text{opposite senses} \end{Bmatrix}$$

The quantity $p = \dfrac{\Sigma(xy)}{n}$ is therefore a sensitive measure of the amount of correlation between X and Y. It is, however, a concrete

quantity, whilst our required coefficient should be abstract We therefore standardize p in terms of X and Y by writing

$$r = \frac{p}{\sigma_x \times \sigma_y} = \frac{\Sigma(xy)}{n\sigma_x\sigma_y} \qquad (1)$$

where σ_x and σ_y are the standardizing factors

The quantity r is known as the coefficient of correlation

It is easy to show that the limits of r as given by this expression are $+1$ and -1 and since the mean product p is extremely sensitive to the respective variations of X and Y r furnishes a satisfactory measure of the degree of interdependence between the two variables

Now write the equation

$$y = b_1 x \qquad (2)$$

where $$b_1 = r\frac{\sigma_y}{\sigma_x} = \frac{p}{\sigma_x^2}$$

and y now represents an estimated value

It can be shown that the straight line represented by this equation is the line of best fit after which we are seeking

This equation may be transferred to its origin by writing

$$Y - \bar{Y} = b_1(X - \bar{X}) \qquad (3)$$

and we have the expression required

This equation is known as a Regression Equation

To sum up

1 The correlation coefficient r represents the amount of the relationship between X and Y

2 The regression equation

$$y = b_1 x = r\frac{\sigma_y}{\sigma_x}x$$

expresses the most probable value of y associated with a given x

The necessary calculations are given in Table 37

$$\left.\begin{aligned} \Sigma(x) &= 0 = \Sigma(y) \\ \Sigma(x + y)^2 &= \Sigma(x^2) + 2\Sigma(xy) + \Sigma(y^2) \\ 548 &= 138 + 246 + 164 \end{aligned}\right\} \quad \textit{Checks}$$

$$\bar{X} = \frac{300}{10} = 30 \quad \bar{Y} = \frac{250}{10} = 25$$

TABLE 37

CALCULATION OF CORRELATION COEFFICIENT FROM TABLE 36

Husband's Age (X) (1)	Wife's Age (Y) (2)	$x = X - 30$ (3)	$y = Y - 25$ (4)	x^2 (5)	y^2 (6)	xy (7)	Check: $x + y$ (8)	Check: $(x + y)^2$ (9)
Years	Years	Years	Years	Years	Years	Years	Years	Years
23	18	−7	−7	49	49	+49	−14	196
27	20	−3	−5	9	25	+15	−8	64
28	22	−2	−3	4	9	+6	−5	25
28	27	−2	+2	4	4	−4	0	0
29	21	−1	−4	1	16	+4	−5	25
30	29	0	+4	0	16	0	+4	16
31	27	+1	+2	1	4	+2	+3	9
33	29	+3	+4	9	16	+12	+7	49
35	28	+5	+3	25	9	+15	+8	64
36	29	+6	+4	36	16	+24	+10	100
300	250	+15 −15	+19 −19	138	164	+127 −4 +123	32 −32	548

In this example, the calculations are facilitated by the fact that $\overline{X}$ and $\overline{Y}$ fall on round numbers.

$$\sigma_x = \sqrt{\frac{138}{10}} = 3{\cdot}715 \text{ yrs.} \quad \sigma_y = \sqrt{\frac{164}{10}} = 4{\cdot}050 \text{ yrs.}$$

$$p = \frac{\Sigma(xy)}{n} = \frac{123}{10} = 12{\cdot}3$$

$$r = \frac{p}{\sigma_x \times \sigma_y} = \frac{12{\cdot}3}{3{\cdot}715 \times 4{\cdot}050} = +\,0{\cdot}8175$$

$$b_1 = r\frac{\sigma_y}{\sigma_x} = \frac{p}{\sigma_x^2}$$

$$= \frac{12{\cdot}3}{13{\cdot}8} = 0{\cdot}8913$$

Therefore $\quad y = 0{\cdot}8913x$

And $\quad Y - 25 = 0{\cdot}8913\,(X - 30)$

i.e. $\quad Y = 0{\cdot}8913X - 1{\cdot}7390$

This last expression represents the regression equation giving the straight line of best fit of Y to X.

Substituting for

$$X = 23 \text{ and } X = 36$$

We have $Y_{23} = 20{\cdot}500 - 1{\cdot}739 = 18{\cdot}761$

$$Y_{36} = 32{\cdot}087 - 1{\cdot}739 = 30{\cdot}348$$

Plotting these values upon the scatter-diagram we have the result shown in Fig 37 It will be noticed that the regression line

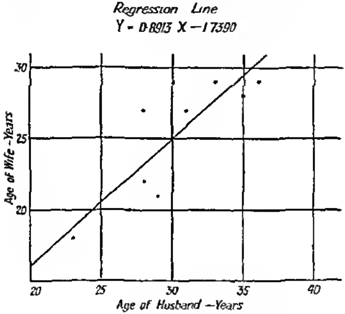


FIG 37

passes through the means of the two variables $\bar{X} = 30$ and $\bar{Y} = 25$

This regression line has been drawn so as to minimize the sum of the squares of the residuals, i e

$$\Sigma\{(0{\cdot}8913X - 1{\cdot}7390) - Y\}^2 = \Sigma(v^2) \text{ is minimized} \qquad (4)$$

There is another Regression Equation relating X to Y This has been omitted in order to prevent confusion at this stage

Tests of Agreement

The expression $\frac{\Sigma(v^2)}{n} = S^2_v$ supplies a test of the agreement between computed and actual values of Y. S^2_v is the variance and S_v the standard deviation of Y calculated not round a fixed point, but round the theoretical value—

$$Y = 0{\cdot}8913X - 1{\cdot}7390$$

The necessary calculations are given in Table 38.

TABLE 38

CORRELATION BETWEEN AGES OF HUSBAND AND WIFE—CALCULATION OF RESIDUALS

Husband's Age (X)	Wife's Age (Y)		Residual (v)	Square of Residual (v^2)
	Computed	Observed		
(1)	(2)	(3)	(4)	(5)
Years	Years	Years		
23	18·76	18	− 0·76	0·5776
27	22·33	20	− 2·32	5·3824
28	23·22	22	− 1·22	1·4884
28	23·22	27	+ 3·78	14·2884
29	24·11	21	− 3·11	9·6721
30	25	29	+ 4·00	16·0000
31	25·89	27	+ 1·11	1·2321
33	27·67	29	+ 1·33	1·7689
35	29·46	28	− 1·46	2·1316
36	30·35	29	− 1·35	1·8225
			+ 10·22 − 10·22	54·3640

Writing $\frac{\Sigma(v^2)^*}{n} = S^2_v$ we have

$$S^2_v = 5{\cdot}4364 \text{ and } S_v = 2{\cdot}3316 \text{ yrs.}$$

The quantity S_v evidently represents the standard deviation of the Y's around the regression line $Y = 0{\cdot}8913X - 1{\cdot}7390$.

The smaller the quantity S_v (standard deviation of the residuals) as compared with σ_v (standard deviation of the original observations) the higher the degree of correlation.

It is easy to prove that

$$S_v = \sigma_v \sqrt{1 - r^2} \quad . \quad . \quad . \quad . \quad (5)$$

* Since $\Sigma(v) = 0$, no correction is required.

quadrants bear positive signs and those in the other two quadrants negative signs The numbers are enclosed in brackets in order to distinguish them from the frequencies

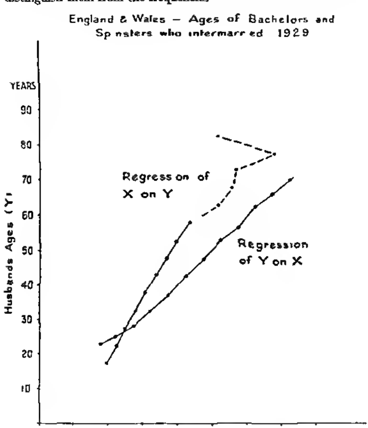


FIG 38

For example there are 458 couples corresponding with a central age of 42 5 years (8 5 units) for the wife and 37 5 years (7 5 units) for the husband Measuring from the trial averages we have $x' = 4$ $y = 2$ $x'y = 8$

TABLE 42

HUSBANDS AGES—MEAN AND STANDARD DEVIATION

Central Age	D stance from 27 5 yr	Frequency	Col (3) × Col (2)	Col (4) × Col (2)
(1)	(2)	(3)	(4)	(5)
Years	Un ts			
17 5	2	5 018	10 036	20 072
22 5	1	106 795	106 795	106 795
27 5	0	115 126	116 831	126 867
52 5	1	52 031	32 031	52 031
57 5	2	10 873	21 746	43 492
42 5	3	4 741	14 223	42 669
47 5	4	2 289	9 156	36 624
52 5	5	1 115	5 565	27 825
57 5	6	535	3 210	19 260
62 5	7	225	1 575	11 023
67 5	8	151	1 208	9 664
72 3	9	40	560	5 240
77 3	10	8	80	800
82 5	11	3	33	363
		78 948	89 187	226 993
			116 831	126 867
			27 644	553 860

$$\bar{Y} = 5\ 5 \quad \frac{27\ 644}{278\ 948}$$

$= 5\ 5 \quad 0\ 099101 = 5\ 400899$ units

$= 27\ 004495$ years

$$= \sqrt{\frac{353\ 860}{278\ 948} - (0\ 099101)^2}$$

$= 1\ 121932$ units

$5\ 60966$ years

TABLE 43

Calculation of Mean Product of Wives' and Husbands' Ages

$x'y'$ (1)	(2)		$fx'y'$ (3)	$x'y'$ (1)	(2)		$fx'y'$ (3)
Units				Units			
110	1	1	110	20	217 + 286	503	10,060
100	5	5	500	18	89 + 29	118	2,124
90	2 + 2	4	360	16	5 + 430	435	6,960
88	1	1	88	15	204 + 196	400	6,000
81	13	13	1,053	14	12 + 6	18	252
80	9	9	720	12	64 + 527 + 693 + 15	1,299	15,588
72	4 + 28	32	2,304	10	208 + 83	291	2,910
70	3	3	210	9	1 + 1135	1,136	10,224
64	33	33	2,112	8	3 + 466 + 458	927	7,416
63	5 + 16	21	1,323	7	7 + 2	9	63
60	1	1	60	6	37 + 1233 + 2012 + 5	3,287	19,722
56	20 + 47	67	3,752	5	97 + 50	147	735
54	5 + 6	11	594	4	341 + 3228 + 239	3,808	15,232
50	1	1	50	3	979 + 1559	2,538	7,614
49	42	42	2,058	2	3493 + 7799 + 3151	14,443	28,886
48	23 + 13	36	1,728	1	14029 + 18124	32,153	32,153
45	5 + 3	8	360				209,585
42	31 + 64	95	3,990	0	115126 + 139960 − 54784	200,302	
40	10 + 8	18	720	− 1	486 + 13997	14,483	− 14,483
36	3 + 78	81	2,916	− 2	91 + 1033 + 120	1,244	− 2,488
35	27 + 36	63	2,205	− 3	34 + 157	191	− 573
32	13 + 4	17	544	− 4	12 + 17 + 5	34	− 136
30	80 + 112	192	5,760	− 5	7 + 5	12	− 60
28	23 + 18	41	1,148	− 6	1 + 3	4	− 24
27	1 + 1	2	54	− 8	1	1	− 8
25	179	179	4,475			278,948	− 17,772
24	7 + 90 + 64	161	3,864				+ 209,585
21	17 + 11	28	588				191,813

This table shows the mean product ($x'y'$) of the distances of each cell centre from the trial averages; the frequency in the cell, and the product of the two. For instance, there are 3 cells with mean product distance = 24, viz.—

Husband's Age (1)	Distance from Trial Average (2)	Wife's Age (3)	Distance from Trial Average (4)	Frequency (5)
Years	Units	Years	Units	
65–70	8	35–40	3	7
55–60	6	40–45	4	90
45–50	4	50–55	6	64
				161

It will be seen that the indications of the two equations are not consistent. This is because in the one case we are making the best estimate of the average age of the wife from the exact age of the husband, and in the other case we are making the best estimate of the average age of the husband from the exact age of the wife.

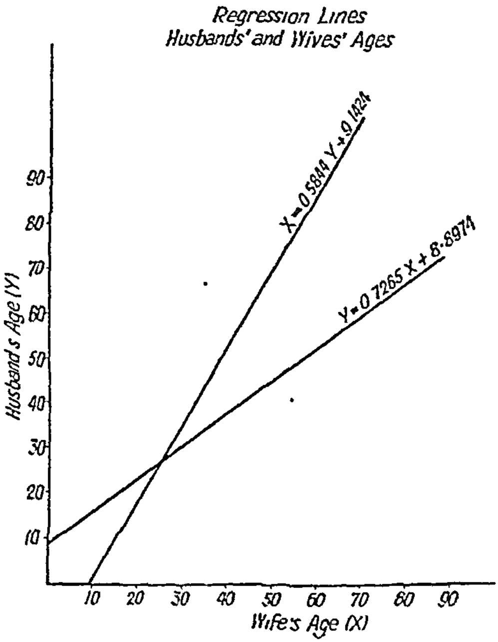


FIG. 39

Significance of the Correlation Coefficient.

The Correlation Coefficient like other Statistical Quantities is subject to Errors of Sampling. It reflects the peculiarities of the sample from which it is drawn, and may not therefore be typical of the population.

It has been customary to define the probable error of the correlation coefficient by the expression—

$$\text{p.e. } (r) = 0{\cdot}6745 \frac{1 - r^2}{\sqrt{n}} \quad . \quad . \quad . \quad . \quad (7)$$

The use of this formula for general purposes is now discredited, and it is not clear why textbook writers nevertheless persist in repeating it.

A simpler method is to test whether r differs significantly from zero [1] The test is—

$$r > \frac{3}{\sqrt{n}} \tag{8}$$

provided n is large

For the data of Table 39 we have—

$$\frac{3}{\sqrt{50}} = 0\ 42$$

This value is greater than that of the coefficient as computed and the correlation is not significant

For the data of Table 40 we have—

$$\frac{3}{\sqrt{278948}} = 0\ 006$$

and the correlation is significant

Correlation of Time Series

If two time series are correlated by the straightforward rule the resulting coefficient will reflect both long term movements (trends) and short term movements It is usually the latter in which we are interested and the traditional method is to correlate the devia tions from moving averages instead of the figures themselves This process involves many theoretical difficulties (especially if the figures are cyclic) and is not recommended for elementary students

General Remarks on the Correlation Co-efficient

The pure theory of correlation is based upon the hypothesis that the data represent a random sample from a homogeneous normal population In practice the condition of normality may be re laxed provided the other conditions are substantially observed Failure of the conditions impairs the validity of the co efficient While it still supplies the best estimate of the correlation for that particular set of data it cannot be used inductively as a basis for general laws Without laying down general rules there is a pre sumption that the method of correlation will be useful in connec tion with controlled experiments or observations under stable con ditions and not otherwise This criterion admits it to the fields of

[1] Consult Fisher s *Statistical Methods for Research Workers* or Tippett s *Methods of Statistics*

agriculture, biology, industrial production, vital statistics, etc., and excludes it for the most part from the fields of business, economics, finance, etc. There is no reason for undertaking long and tedious calculations unless they are likely to lead to results of permanent interest. The ordinary textbook examples, purporting to correlate Demand with Supply, etc., are useless for practical purposes.

Other Methods.

Other and simpler methods of measuring correlation will be found in the textbooks.[1] They are not usually very satisfactory.

[1] *Cf.* King's *Elements of Statistical Method* (Chapter XVII) or Boddington's *Statistics and Their Application to Commerce* (Chapter X).

CHAPTER XVI

INDEX NUMBERS

In this Chapter it is proposed to discuss the theory of Index Numbers and the general principles governing their construction The application of these principles to concrete cases will be postponed to Part II

An Index Number is a device for estimating the relative movements of a Statistical Variate in cases where measurement of its actual movements is inconvenient or impossible The method of index numbers is particularly appropriate when the variate in question is unstable in its composition or when the labour and expense of direct observation would be prohibitive Changes in price levels afford a good illustration of the type of problem involved No records are available of the actual volumes and prices of all commodities that change hands but market quotations can be obtained for the principal articles of trade and the significance of each may be estimated with fair accuracy Business prosperity is the resultant of the interactions of a highly complex system of forces and there is no definite standard whereby it can be measured Yet by studying certain phases of business activity which are peculiarly symptomatic of the rest it is possible to construct an *index* of business prosperity whose movements afford general indications of the trend of activity without pretensions to accurate measurement

Relatives

The basis of the method of Index Numbers is the Relative

A relative is simply the ratio between the price of the current year[1] and that of a standard year and is usually expressed as a percentage of the latter In this case the figure 100 is known as the base, and the standard year as the base year

A relative series is formed from the corresponding absolute series by multiplication by a constant factor

[1] In order to fix our ideas the theory will be discussed in terms of *prices* and *years* Application to quantities other than prices and periods other than years is simple

In Table 44, col. (5), the actual prices of copper in £ per ton are given as 75, 68, 69⅓, $63\frac{3}{16}$, and 54¾. Multiplication by $1{\cdot}\dot{3}$ converts this series into 100, 90·67, 92·44, 84·25, and 73·00.

An Index Number as a Combination of Relatives.

A combination of these relatives is known as an Index Number. The combination is effected by aggregation or averaging. In theory any form of average may be employed, and the latter may be either simple or weighted.

In the following sections P_{01} represents the index number for the current year with respect to the base year—

$p_0'\ p_0''\ \ldots\ p_0^n$ represent the prices of the base year; and

$p_1'\ p_1''\ \ldots\ p_1^n$ represent the prices of the current year.

The dashes refer to the commodities and the subscripts to the years. In order to avoid confusion, the external multiplier (100) has been omitted.

Specimen Formulae.

A simple arithmetic average of relatives would be written

$$P_{01} = \frac{1}{n}\left\{\frac{p_1'}{p_0'} + \frac{p_1''}{p_0''} + \ldots + \frac{p_1^n}{p_0^n}\right\} = \frac{1}{n}\,\Sigma\left(\frac{p_1}{p_0}\right) \quad . \quad . \quad (1)$$

And a simple geometric average

$$P_{01} = \sqrt[n]{\frac{p_1'}{p_0'} \times \frac{p_1''}{p_0''} \times \ldots \frac{p_1^n}{p_0^n}} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

And a weighted arithmetic average

$$P_{01} = \frac{1}{\Sigma w}\left\{w'\frac{p_1'}{p'_0} + w''\frac{p_1''}{p_0''} \quad . + w^n\frac{p_1^n}{p_0^n}\right\}$$

$$= \frac{1}{\Sigma w}\,\Sigma\left(w\frac{p_1}{p_0}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The *Statist* Index of Minerals.

The following Table illustrates the calculation of the *Statist* index of *Wholesale Prices of Minerals* for the years 1913, 1921, 1924, and 1930

TABLE 44

CALCULATION OF *Statist* INDEX OF WHOLESALE PRICES OF MINERALS 1913 1921 1924 AND 1930

AVERAGE PRICES OF COMMODITIES—MINERALS

Year	Iron			Copper Standard	Tin Straits	Lead English Pig	Coal		Minerals Total	Index No of Minerals = Col (10 ÷ …)
	Scottish Pig	Cleveland (Middlesbrough Pig)	Bars Common				Wallsend Hetton in London	Average Export Price		
	Shillings and Pence per Ton	Shillings and Pence per Ton	£ per Ton	£ per Ton	£ per Ton	£ per Ton	Shillings per Ton	Shillings and dec per Ton		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Average 1867–77	69	60	8½	75	105	20⅛	22	12.5		
1913	65 6	58 3	7¾	68	201	19⅛	21½	13.94		
1921	168 6	137 4	19⅛	69⅝	171	24	32½	34.83		
1924	96 8	88 2	12⅜	63 ¼	251	35⅛	27½	23.38		
1930	76	67	9⅓	54⅜	144 ⅞	19⅛	24 ½	10.64		

INDEX NUMBERS OR PERCENTAGES (Average 1867–77 = 100)[2]

Year	Iron (2)–(3)	Bars Common (4)	Copper (5)	Tin (6)	Lead (7)	Wallsend (8)	Export Price (9)	Minerals Total (10)	Index No (11)
Average 1867–77	100	100	100	100	100	100	100	00	100
1913	96.00	93.94	90.67	191.43	93.29	97.73	111.52	774.58	110.65
1921	236.54	231.82	92.44	162.86	118.05	146.59	278.64	1266.94	180.99
1924	143.52	151.52	84.25	239.05	174.69	125.00	187.04	1105.07	157.87
1930	110.91	120.84	73.00	138.01	95.12	112.04	133.12	783.64	111.95

[1] Wholesale Prices of Commodities in 1930 *J R S S* Vol XCIV (1931) p 280

[2] The *Statist* calculates to the nearest whole number This is sufficiently accurate for practical purposes but for purposes of illustration two places of decimals have been retained

The *Statist* index is a simple arithmetic average of price relatives based on the average prices of the years 1867–77 (Formula (1)).

Column (11) shows the movements in the average wholesale price of the mineral products included in the index number, upon the assumption that such movements may be adequately represented by a simple arithmetic average.

Technique of Index Number Construction.

The technique of Index Number construction involves four major problems—

1. The choice of items to be included.
2. The base period.
3. The form of average to be adopted.
4. The weighting system.

These matters will be discussed with especial reference to commodity prices, which afford the most natural approach to the problem. The principles involved will, of course, apply to any type of index number.

Choice of Items.

Since it is impossible to include all items within the field of inquiry, the items selected should be representative of the tastes, habits, or requirements of the class of purchaser concerned. They should also be easily identifiable and unlikely to vary in quality. These restrictions narrow the field of choice considerably. Reliable quotations are available for food-stuffs, raw materials of industry, and semi-manufactured articles. Manufactured articles are more difficult to deal with because of variations in quality, and personal services still more difficult, since they are not embodied in tangible goods, and there is no standard (apart from the monetary standard) whereby to measure them.

On this account the more widely used price index numbers are confined to wholesale prices of food-stuffs, raw materials, and semi-manufactured articles, whose qualities and descriptions are standardized. They are therefore producers' index numbers. Comparatively little progress has been made with consumers' index numbers. Most countries publish official indices of retail prices or cost of living for the working classes, but there are no satisfactory consumers' indices for other classes of the community.

Practice varies with regard to the number of items included The larger the number the nearer may the index attain towards the conditions of random sampling [1] and the greater the tendency of errors to compensate one another On the other hand the inclusion of a large number of quotations involves difficulty expense and delay

The *Board of Trade Wholesale Price Index* includes 200 items and *The Economist Wholesale Price Index* 58

With these must be contrasted the so-called sensitive indices based upon a small number of items (say 20) supposed to be especially sensitive [2] to changes in business conditions

Quotations are usually obtained from trade journals or from leading firms of dealers in the commodities in question In some cases quotations are based upon average import or export prices

Occasionally an article becomes temporarily unobtainable or obtainable only at a prohibitive price Since under these circumstances the article would not be bought to any appreciable extent logic demands that it be eliminated from the index The elimination of an article usually involves a considerable amount of recalculation and in order to avoid this difficulty it is usual to insert nominal quotations until normal conditions have reasserted themselves

Should it become definitely necessary to introduce a new item or drop an old one one must calculate a new series and splice it on to the old series The splicing is effected by calculating the index on both systems (the old and the new) for the transition year and then multiplying the new series by a factor chosen so as to equate the two series at that point

Questions of some delicacy are raised when the price of an article is controlled by the Government but illicit dealings take place at uncontrolled prices So far as Britain is concerned this uestion is academic but it possesses considerable practical interest regards countries like Russia.

CALCULATION OF

oice of Base

'here are two methods available

a) Fixed base

b) Chain base

[1] ee Chapter XIV

[2] f the index of fifteen pr mary commodities compiled by the *Bank of* *nd*

TABLE 45

RECALCULATION OF *Statist* INDEX OF MINERAL PRICES UPON DIFFERENT BASES

RELATIVES OF MINERAL PRICES

Year (1)	Iron: Scottish Pig (2)	Iron: Cleveland (Middlesbrough Pig) (3)	Iron: Bars Common (4)	Copper Standard (5)	Tin Straits (6)	Lead English Pig (7)	Coal Wallsend Hetton in London (8)	Average Export Price (9)	Minerals Total (10)	Index No. of Minerals = Col. (10) ÷ 7 (11)
				1913 = 100						
1913	100		100	100	100	100	100	100	700	100
1921	246·40		246·77	101·95	85·08	126·54	150·00	249·86	1206·60	172·37
1924	149·50		161·29	92·92	124·86	187·26	127·90	167·72	1011·45	144·49
1930	115·53		128·04	80·51	72·09	101·96	115·26	119·37	733·36	104·77
				1921 = 100						
1913	40·59		40·52	98·08	117·54	79·02	66·67	40·02	482·44	68·92
1921	100		100	100	100	100	100	100	700	100
1924	60·67		65·36	91·14	146·78	147·98	85·27	60·19	657·39	93·91
1930	46·89		52·13	78·97	84·74	80·58	76·84	42·84	462·99	66·14
				1924 = 100						
1913	66·89		62·00	107·62	80·08	52·83	78·18	59·62	507·22	72·46
1921	164·81		153·00	109·72	68·13	67·58	117·27	148·97	829·48	118·50
1924	100		100	100	100	100	100	100	700	100
1930	77·28		79·75	86·55	57·73	54·45	90·11	71·17	517·14	73·88
				1930 = 100						
1913	86·56		77·74	124·21	138·71	98·08	86·76	83·85	695·91	99·42
1921	213·27		191·18	126·36	118·00	124·11	130·14	209·31	1112·37	158·91
1924	129·40		125·39	115·41	173·21	183·65	110·97	140·50	978·53	139·79
1930	100		100	100	100	100	100	100	700	100

Fixed Base Method

With the Fixed Base Method a definite year (or average of a period of years) is chosen and adhered to for an indefinite time The period selected should be a period of normal conditions and free from disturbances likely to affect the index

The *Statist* index number is based upon average prices for the period 1867–77 [1]

The method of the simple arithmetic average of relatives [2] gives rise to different results according to the base year chosen The whole series differs not only in its absolute values (this is immaterial) but also in its relative movements

In order to illustrate this fact Table 45 has been drawn up to show the results of re-calculating the *Statist Index Number of Minerals* upon four alternative bases viz 1913 1921 1924 and 1930

Table 46 summarizes and compares these results with the results given in Table 44

TABLE 46

Statist INDEX OF MINERAL PRICES RE-CALCULATED ON VARIOUS BASES

Year	Base				
	Average 1867 77	1913	1921	1924	1930
(1)	(2)	(3)	(4)	(5)	(6)
1913	110 65 *100 00*	100 00 *100 00*	68 92 *100 00*	72 46 *100 00*	99 42 *100 00*
19 1	180 99 *163 57*	172 37 *172 37*	100·00 *145 10*	118 50 *163 54*	158 91 *159 84*
1924	157 87 *142 65*	144 49 *144 49*	93·91 *136 26*	100 00 *138 01*	139 79 *140 60*
1930	1 95 *101 17*	104 77 *104 77*	66 14 *95 97*	73 88 *101 96*	100 00 *100 58*

The roman figures give the indices as actually calculated in Tables 44 and 45 the base year being equated to 100 in each case Figures presented on this plan are however difficult to compare and the *italic* figures show the results of equating all the indices to 100 in the year 1913

This last operation of equating the year 1913 does not involve

[1] See page 159
[2] Form I1 (1) p 157

change of weights base.[1] Since an index involves a relative series, only its proportional movements are significant. The absolute values are immaterial.

The differences in the indications of the five series will be noted. The most erratic figures are those in the row and column headed "1921."

Upon analysis it will be seen that the differences are really due to differences of weighting. Although this purports to be a simple (unweighted) index, it is really an arbitrarily weighted index, the arbitrary element being involved in the choice of base year. If in the base year the price of any given commodity is abnormal, it will exercise an influence to correspond upon the total. The argument may be crystallized in the statement that change of base year is equivalent to change of weights.

Chain Base Method.

With the Chain Base Method, each year is calculated upon the preceding year as base, and the results are chained together afterwards.

Applying the notion to an index based on the simple arithmetic average, the operation is as follows—

$$P_{01} = \frac{1}{n}\,\Sigma\left(\frac{p_1}{p_0}\right)$$

$$P_{12} = \frac{1}{n}\,\Sigma\left(\frac{p_2}{p_1}\right)$$

$$P_{23} = \frac{1}{n}\,\Sigma\left(\frac{p_3}{p_2}\right) \text{ etc.} \quad \ldots \quad (4)$$

Whence chaining the results together

$$P'_{01} = P_{01}$$

$$P'_{02} = P_{01} \times P_{12}$$

$$P'_{03} = P_{01} \times P_{12} \times P_{23}, \text{ etc.} \quad \ldots \quad (5)$$

$$P'_{02} = \frac{1}{n^2}\left\{\frac{p_1'}{p_0'} + \frac{p_1''}{p_0''} + \ldots \frac{p_1^n}{p_0^n}\right\}\left\{\frac{p_2'}{p_1'} + \frac{p_2''}{p_1''} + \ldots \frac{p_2^n}{p_1^n}\right\} \quad (6)$$

Whilst

$$P_{02} = \frac{1}{n}\left\{\frac{p_2'}{p_0'} + \frac{p_2''}{p_0''} + \ldots \frac{p_2^n}{p_0^n}\right\}. \quad \ldots \quad (7)$$

[1] See the further discussion on p. 175.

Table 47 shows the results of applying the Chain Base system to the *Statist* index of *Sugar Tea and Coffee* The roman type denotes the relatives as recorded by the *Statist* whilst the *italic figures show the result of basing each relative upon its predecessor* The results are totalled in column (6) and averaged in column (7) Column (8) shows the result of chaining e g —

$$83 \times \frac{100.8}{100} = 83.7$$

$$83.7 \times \frac{134.5}{100} = 112.5 \text{ etc}$$

TABLE 47

SUGAR TEA AND COFFEE PRICES—CALCULATION OF CHAIN BASE INDEX

Year	Sugar (I)	Sugar (II)	Coffee	Tea	Total	Average	Chain Index
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1921	81	77	119	55	332	83.0	85.0
1922	62	54	128	82	326	81.5	85.7
	76	*70*	*108*	*149*	*403*	*100.8*	
1923	104	87	111	100	402	100.5	112.5
	168	*161*	*87*	*122*	*538*	*134.5*	
1924	93	75	154	96	418	104.5	115.3
	89	*86*	*139*	*96*	*410*	*102.5*	
1925	60	43	165	88	356	89.0	92.8
	65	*57*	*107*	*93*	*322*	*80.5*	
1926	60	44	159	89	352	88.0	100.0
	100	*103*	*96*	*101*	*400*	*100.0*	
1927	62	47	139	84	332	83.0	90.8
	103	*107*	*87*	*94*	*391*	*97.8*	
1928	5	40	146	77	314	78.5	82.6
	82	*85*	*105*	*92*	*364*	*91.0*	
1929	42	31	140	74	287	71.8	72.6
	82	*77*	*96*	*96*	*351*	*87.8*	
1930	31	22	95	67	215	53.8	55.0
	74	*71*	*68*	*90*	*303*	*75.8*	

The differences between the results of the two systems may be ascribed to the small number of commodities included and to the erratic movements of individuals

Advantages and Disadvantages of Chain Base Method

The advantages claimed for the chain base method are that it provides a direct comparison between each year and the next

which is more interesting to commercial users than indirect comparison through the medium of a possibly remote base year, and that it facilitates the introduction of new items and the dropping of old ones.

Its disadvantage[1] lies in the fact that it usually possesses a definite bias which vitiates comparisons between distant periods.

The Form of Average.

In theory it is possible to employ any form of average. For practical purposes the choice lies between the arithmetic and the geometric mean. The median and mode are usually erratic, and other forms of average are complicated or otherwise unsuitable.

The arithmetic mean has the advantage of intelligibility, and when weighted by the quantity of the commodities purchased the result corresponds with a definite objective quantity, viz. the total outlay on the commodities in question. On the other hand, the arithmetic mean suffers from a bias which it is difficult to remove.

Reversibility.

The Arithmetic Average of Relatives is Not Reversible. The result of calculating the current year upon the base year does not agree with the result of calculating the base year upon the current year. To ensure consistency we should have

$$P_{01} \times P_{10} = 1$$

In other words, the Index for the current year upon the base year, and the Index for the base year upon the current year should be reciprocal to each other.

With the arithmetic average of relatives (Formula (1))

$$P_{01} \times P_{10} > 1$$

To test this proposition let us refer to Table 46.

The index of minerals for 1921 with base 1913 is 172·37 and $P_{01} = 1{\cdot}7237$.

The index for 1913 with base 1921 is 68·92, and $P_{10} = 0{\cdot}6892$.

$$\begin{aligned} P_{01} \times P_{10} &= 1{\cdot}7237 \times 0{\cdot}6892 \\ &= 1{\cdot}1880 \end{aligned}$$

[1] This disability may be avoided by the employment of the geometric mean (see page 166).

The algebraic proof of this proposition is simple

$$P_{01} = \frac{1}{n}\left\{\frac{p_1}{p_0} + \frac{p_1}{p_0} + \quad \right\} \text{ whereas}$$

$$\frac{1}{P_{10}} - \frac{1}{\frac{1}{n}\left\{\frac{p_0}{p_1} + \frac{p_0}{p_1} + \quad \right\}} \tag{8}$$

and these quantities are not reciprocal

Geometric Mean

The simple geometric mean is reversible

$$\text{For } P_{01} - \sqrt[n]{\frac{p_1}{p_0} \times \frac{p_1}{p_0} \qquad \frac{p_1^n}{p_0^n}}$$

$$= \frac{1}{\sqrt[n]{\frac{p_0}{p_1} \times \frac{p_0}{p_1} \quad \times \frac{p_0}{p_1^n}}} \quad \frac{1}{P_{10}} \tag{9}$$

With the simple geometric mean the fixed base and chain base methods agree

$$P_0 \;-\sqrt[n]{\frac{p_1}{p_0} \times \frac{p_1}{p_0} \qquad \times \frac{p_1^n}{p_0^n}} \times \sqrt[n]{\frac{p_2}{p_1} \times \frac{p_2}{p_1} \qquad \times \frac{p_2^n}{p^n_1}}$$

$$- \sqrt[n]{\frac{p_2}{p_0} \times \frac{p_2}{p_0} \qquad \times \frac{p_2^n}{p_0^n}} \tag{10}$$

The above theorems also apply to the weighted geometric mean provided the weights are kept constant [1]

It is also claimed for the geometric mean that it is the natural type to use when it is a question of ratios and that it is especially suitable for quantities (like index numbers) which have a fixed lower limit (zero) but no assignable upper limit

On the other hand the geometric mean is less intelligible and more difficult to calculate and the application of weights does not lead to an objective result

The geometric mean is necessarily less than the arithmetic mean [2]

[1] This proposition does not apply to a geometric mean with variable weights i e weights that change according to the years in question

[2] Unless all the items are equal in which case the two means are equal

This fact constitutes an argument in favour of the former, for when the price of an article rises there is a tendency for less to be purchased, and for the weight of that article (or its contribution to the average) to be reduced. The fact that the geometric mean weights large items less than the arithmetic therefore provides a corrective tendency.

On balance, the advantages of the geometric mean preponderate, and in this country (at any rate) it is now preferred by the leading authorities.

Table 48, on p. 168, illustrates the calculation of the simple geometric mean for the *Statist* index number of minerals.

For 1913 the arithmetic mean of the relatives in columns (2) to (9) was 110·6. The arithmetic mean of their logarithms was 2·0288, of which the anti-logarithm is 106·9. The figure 106·9 is therefore the geometric mean of the price relatives in columns (2) to (9).

The Weighting System.

So far, no special assumptions have been made regarding weights. The use of the simple average implies that the relatives concerned are, actually or notionally, equivalent in importance. Certain classes of problem, such as the price level for a generalized purchaser or the stock and share level for a generalized operator, are so indefinite in their terms that it seems invidious to assign any particular item precedence over any other. But in most classes of inquiry there are some grounds for differentiation, and expression should, if possible, be given to this fact by means of an appropriate weighting system.

Since the influence exercised by a movement in a given item depends not only upon the essential fact of the movement itself, but also upon the accident of its price level at the moment, a simple index number is in reality an arbitrarily weighted average.

Statistics relating to quantities of goods produced, sold, and consumed, are more plentiful than formerly, and there is less excuse for the argument that index numbers cannot be weighted because of absence of information.

It is not necessary to seek after a high degree of accuracy in weighting, since it may be proved that errors in weights have little effect compared with variations in prices.

The problem of the weighting of index numbers is still in the

TABLE 48

Statist Index of Mineral Prices. Recalculation by Geometric Mean

(The italic figures are the logarithms of the respective relatives, except in Column (11), where they represent the antilogarithms of Column (10).)

Year	Iron: Scottish Pig	Iron: Cleveland (Middlesbrough Pig)	Iron: Bars Common	Copper Standard	Tin Straits	Lead English Pig	Coal Wallsend Hetton in London	Average Export Price	Average of Logs	Index No
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1913	96.00 *1.9823*		93.84 *1.9723*	90.67 *1.9574*	191.43 *2.2819*	93.20 *1.9693*	97.73 *1.9900*	111.52 *2.0473*	*2.0258*	110.6 *106.1*
1921	236.54 *2.3739*		231.82 *2.3651*	92.44 *1.9659*	162.80 *2.2118*	118.05 *2.0721*	146.50 *2.1662*	278.64 *2.4449*	*2.2298*	181.0 *169.7*
1924	143.52 *2.1569*		151.52 *2.1804*	84.25 *1.9256*	237.05 *2.3758*	174.60 *2.2422*	125.00 *2.0969*	187.04 *2.2719*	*2.1789*	157.9 *151.0*
1930	110.31 *2.0448*		120.84 *2.0820*	73.01 *1.8633*	138.02 *2.1399*	95.12 *1.9783*	112.04 *2.0494*	133.12 *2.1243*	*2.0466*	112.0 *111.4*

controversial stage, and it will be impossible in the space available to do more than indicate its main features.

Methods of Weighting.

There are two methods of weighting price indices—

1. Applying the weights to the prices themselves. This leads to the Aggregative index number.

2. Applying the weights to the price relatives. This method leads to the Average of ratios.

Aggregative Method.

Since total value = price × quantity, the appropriate weights to apply to actual prices are physical quantities. The latter are applied direct to the prices and there is no need to calculate the price relatives.

The quantities employed may be—

1. The actual quantities of the base year.
2. The actual quantities of the current year.
3. Fixed weights or estimated quantities having reference to a supposed typical year.

Symbolically, these may be written—

Base Year Weighting

$$P_{01} = \frac{q_0'p_1' + q_0''p_1'' + \ldots + q_0^n p_1^n}{q_0'p_0' + q_0''p_0'' + \ldots + q_0^n p_0^n} = \frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} \quad . \quad (11)$$

$$= \frac{\text{Base year quantities at current prices}}{\text{Base year quantities at base prices}}$$

Current Year Weighting

$$P_{01} = \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$= \frac{\text{Current year quantities at current prices}}{\text{Current year quantities at base prices}}$$

Fixed Weights

$$P_{01} = \frac{\Sigma w p_1}{\Sigma w p_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Where q_0 represents the actual quantity of the base year

q_1 current year

w a fixed weight supposed to depend upon a *typical* year

These formulae are suitable for use with either the fixed or the chain base method With formula (13) the weights are absolutely fixed irrespective of the years in question With formula (11) the weights are fixed so long as the same base year is retained and with formula (12) they vary irrespectively of the base

Since there is no logical reason for preferring the weights of the base year to those of the current year or *vice versa* certain mixed formulae have been proposed of which the best known are—

Crossed Weight Formula (Fisher s Ideal Formula)—

$$P_{01} = \sqrt{\frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}} \tag{14}$$

Hybrid Weight Formula—

$$P_{01} = \frac{\Sigma \frac{q_0 + q_1}{2} p_1}{\Sigma \frac{q_0 + q_1}{2} p_0} = \frac{\Sigma (q_0 + q_1) p_1}{\Sigma (q_0 + q_1) p_0} \tag{15}$$

Average of Ratios

In this case the price relatives are fundamental not the prices and we weight by (quantity $\times$ price) $=$ total expenditure or outlay in order to compensate for the price element of the denominator

Applying the same line of argument we write—

Weighting by Base Year Expenditure—

$$P_{01} = \frac{q_0 p_0 \frac{p}{p_0} + q_0 p_0 \frac{p_1}{p_0} + }{q_0 p_0 + q_0 p_0 + } = \frac{\Sigma (q_0 p_0) \frac{p_1}{p_0}}{\Sigma (q_0 p_0)} \tag{16}$$

Here the price relatives $\left(\frac{p_1}{p_0}\right)$ are weighted by total expenditure $(q_0 p_0)$

In this particular form of the weighted price ratio the p's cancel out and the expression becomes

$$\frac{q_0'p' + q_0''p_1'' + \cdots}{q_0'p_0' + q_0''p_0'' + \cdots} = \frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} \quad . \quad . \quad . \quad (17)$$

which is identical with aggregative No. (11).

Weighting by Current Year Expenditure

$$P_{01} = \frac{(q_1'p_1')\dfrac{p_1'}{p_0'} + (q_1''p_1'')\dfrac{p_1''}{p_0''} + \cdots}{(q'p_1') + (q_1''p_1'') + \cdots} \quad . \quad . \quad . \quad (18)$$

In this case there is no cancellation.

It is possible to hybridize the weights.

First Hybrid—

$$P_{01} = \frac{(q_0'p_1')\dfrac{p_1'}{p_0'} + (q_0''p_1'')\dfrac{p_1''}{p_0''} + \cdots}{(q_0'p_1') + (q_0''p_1'') + \cdots} = \frac{\Sigma(q_0p_1)\dfrac{p_1}{p_0}}{\Sigma(q_0p_1)} \quad (19)$$

Again there is no cancellation.

Second Hybrid—

$$P_{01} = \frac{(q_1'p_0')\dfrac{p_1'}{p_0'} + (q_1''p_0'')\dfrac{p_1''}{p_0''} + \cdots}{(q_1'p_0') + (q_1''p_0'') + \cdots} = \frac{\Sigma(q_1p_0)\dfrac{p_1}{p_0}}{\Sigma(q_1p_0)} \quad . \quad (20)$$

In this case there is cancellation and the expression reduces to

$$\frac{\Sigma q_1 p_1}{\Sigma q_1 p_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

which is identical with Formula (12).

Crosses between these formulae can be arranged as before.

It will be seen that the weighted average of ratios (i.e. price relatives weighted by expenditure) admits of more variety of expression than the aggregative form. It must necessarily be employed when the actual prices or other quantities in question are not available, but only estimates of their relative movements. For example, in a study of wage levels it may be impossible to obtain information as to actual wages paid, whilst it may be possible to

give sufficiently accurate estimates of their relative movements In that case it would be necessary to make the best estimates possible of the relative movements of wages in each occupation and weight them by quantities based upon estimated total wage bills

Application of Weighting Methods to Statistics of Crop Prices [1]

Tables 49–53 illustrate the application of these principles to the calculation of an index of prices of crops for the *United States of America*

Table 49 shows the average yearly prices of eight principal crops calculated to the nearest cent per unit and the corresponding production in millions of units

Table 50 shows the results of applying the quantities to the prices Since quantity × price = value columns (2) (7) (9) and (11) check up with the actual values as recorded by the *Department of Agriculture* The remaining columns of course represent notional values

TABLE 51

INDICES OF CROP PRICES UNITED STATES OF AMERICA 1927 29
AGGREGATIVE METHOD

Formula	Index of Crop Prices (Base 19 6)		
	1927	1928	1929
(1)	(2)	(3)	(4)
Base year weighting—			
$100 \times \frac{\Sigma q_o p_x}{\Sigma q_o p_o}$	110 5	105 0	109 5
Current year weighting—			
$100 \times \frac{\Sigma q_x p_x}{\Sigma q_x p_o}$	105 7	101 1	107 4
Crossed weights—			
$100 \times \sqrt{\frac{\Sigma q_o p_x}{\Sigma q_o p_o} \times \frac{\Sigma q_x p_x}{\Sigma q_x p_o}}$	108 1	103·0	108 5

[1] It is difficult at the moment to find reliable statistics of prices and quantities in conjunction but the figures quoted here comply with this condition The object of this section is *to illustrate principles* which are likely to be removed from the academic to the practical sphere during the next few years

Explanation—

$$100 \times \frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} = \frac{691419{\cdot}1}{625567{\cdot}6} = 110{\cdot}5$$

$$100 \times \frac{\Sigma q_2 p_2}{\Sigma q_2 p_0} = \frac{676397{\cdot}1}{669189{\cdot}4} = 101{\cdot}1$$

For the price factor to be reversible, the index for the base year on the current year must be the reciprocal of the index of the current year upon the base year. Let us test the formulae—

$$\frac{\Sigma q_0 p_1}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_0}{q_1 p_1} = 1{\cdot}105 \times \frac{1}{105{\cdot}7} = 1{\cdot}045$$

$$\frac{\Sigma q_2 p_2}{\Sigma q_2 p_0} \times \frac{\Sigma q_0 p_0}{\Sigma q_0 p_2} = 1{\cdot}011 \times \frac{1}{105{\cdot}0} = 0{\cdot}963$$

These indices are therefore not reversible.

The crossed-weight index is reversible.

$$\text{for } \sqrt{\frac{\Sigma q_0 p_x}{\Sigma q_0 p_0} \times \frac{\Sigma q_x p_x}{\Sigma q_x p_0}} \times \sqrt{\frac{\Sigma q_x p_0}{\Sigma q_0 p_x} \times \frac{\Sigma q_0 p_0}{\Sigma q_x p_x}} = 1 \text{ (identically).}$$

The student should test for himself the result of employing the Hybrid weight formula given on page 170.

Table 52 illustrates the calculation of the index by the method of averaging ratios for the year 1927. Column (2) shows the price relatives for 1927 on 1926. Columns (3), (5), (7), and (9) show the various weights that can be employed, and columns (4), (6), (8), and (10) the results of applying these weights.

The results are shown in Table 53, page 177.

There is a useful alternative to the systems of strict weighting described above. Unimportant commodities are assigned one quotation apiece, and more important commodities two or more quotations roughly in proportion to their importance. The result is curious: arithmetically speaking, the index is unweighted, but, statistically speaking, it has the properties of a weighted index. This device is useful and is likely to be extended.

Weights Base and Date Base.

The Weights Base represents the standard year to which the prices of the current year are referred. The Date Base represents the year equated to 100. These two years are not necessarily the

TABLE 52

CALCULATION OF INDICES OF CROP PRICES UNITED STATES OF AMERICA 1927
(Average of Ratios Method)

Crop (1)	$\frac{p_1}{p_0}$ (2)	$\Sigma(q_0p_0)$ (3)	(2) × (3) (4)	$\Sigma(q_1p_1)$ (5)	(2) × (5) (6)	$\Sigma(q_0p_1)$ (7)	(2) × (7) (8)	$\Sigma(q_1p_0)$ (9)	(2) × (9) (10)
Corn	112.6	173	19479.8	200	22520.0	195	21957.0	177	19930.2
Wheat	93.1	100	9310.0	98	9123.8	93	8658.3	105	9775.5
Oats	113.1	50	5655.0	53	5994.3	56	6333.6	47	5315.7
Barley	117.9	11	1296.9	18	2122.2	13	1532.7	15	1768.5
Potatoes	68.2	50	3410.0	39	2659.8	34	2318.8	57	3887.4
Cotton	179.8	98	17620.4	127	22834.6	176	31644.8	71	12765.8
Hay	80.5	121	9740.5	120	9660.0	98	7889.0	149	11994.5
Tobacco	116.5	24	2796.0	26	3029.0	28	3262.0	22	2563.0
TOTALS	881.7	627	69308.6	681	77943.7	695	83596.2	643	68000.6

TABLE 53

INDICES OF CROP PRICES, UNITED STATES, FOR THE YEAR 1927
(Average of Ratios Method)

Formula	Index 1927 (Base 1926)	Remarks
$100 \times \dfrac{\Sigma q_0 p_0 \frac{p_1}{p_0}}{\Sigma q_0 p_0}$	110·5	Same as aggregative formula— $\dfrac{\Sigma q_0 p_1}{\Sigma q_0 p_0}$
$100 \times \dfrac{\Sigma q_1 p_1 \frac{p_1}{p_0}}{\Sigma q_1 p_1}$	114·5	
$100 \times \dfrac{\Sigma q_0 p_1 \frac{p_1}{p_0}}{\Sigma q_0 p_1}$	120·6	
$100 \times \dfrac{\Sigma q_1 p_0 \frac{p_1}{p_0}}{\Sigma q_1 p_0}$	105·7	Same as aggregative formula— $\dfrac{\Sigma q_1 p_1}{\Sigma q_1 p_0}$
For comparison—		
Unweighted arithmetic average .	110·2	
Crossed weight aggregative formula .	108·1	
Unweighted geometric average .	106·1	

same. The difference is best explained by an example. The *Ministry of Labour* Cost of Living Index is based upon 1914 (July). To change the weights base to 1930, it would be necessary to recalculate the whole series, item by item, using weights appropriate to 1930. To change the date base, all that is necessary is to multiply the existing index by a factor calculated to yield a figure of 100 in 1930. In general, the results of the two methods will differ. The first gives virtually a new index, while the second merely gives the old index in disguise.

Summary and Conclusion.

Whilst the differences that emerge from the calculation of the indices by various formulae are sometimes substantial, it should be observed that the indices all, without exception, point in the same direction. Thus an index may be trustworthy in its tendencies without being reliable to the last digit.

The discrepancies referred to are due in large part to the small number of items included. Had the index included fifty times as

many items, more room would have been afforded for the cancellation of unbiased errors, and the results would have been in closer agreement

The form of index number to use must depend upon circumstances, e g the type of problem, the extent and reliability of the information obtainable, the staff available for computing, the possibility of obtaining satisfactory weights and a wide range of really representative prices, etc

The two most promising lines of development are—

A simple index with chain base and geometric mean

An aggregative index of the crossed variety

It follows that the problem of index numbers is not determinate. There is always room for doubt and no index is really significant to within more than a few points During normal times, however, the majority of index numbers calculated upon rational principles show remarkably close agreement and to the extent to which this occurs their indications may be trusted

Various practical problems relating to index numbers will be treated in Part II

CHAPTER XVII

FINITE DIFFERENCES, INTERPOLATION, GRADUATION, AND CURVE FITTING

I. Finite Differences

The calculus of Finite Differences studies the changes that take place in a dependent variable consequent upon finite changes in an independent variable with which it is associated. The method is utilized in connection with the operations of Interpolation and Graduation.

Consider the following Difference Table.

TABLE 54
Difference Table

Argument x (1)	Entry y (2)	Differences Δy (3)	$\Delta^2 y$ (4)	$\Delta^3 y$ (5)	$\Delta^4 y$ (6)
a	y_0				
		Δy_0			
$a+h$	y_1		$\Delta^2 y_0$		
		Δy_1		$\Delta^3 y_0$	
$a+2h$	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$
		Δy_2		$\Delta^3 y_1$	
$a+3h$	y_3		$\Delta^2 y_2$		$\Delta^4 y_1$
		Δy_3		$\Delta^3 y_2$	
$a+4h$	y_4		$\Delta^2 y_3$		
		Δy_4			
$a+5h$	y_5				

Column (1) shows the independent variable (x) which is supposed to advance by equal increments of h. Column (2) shows the corresponding values of the dependent variable $y(=f(x))$, and columns (3) onwards show the successive differences of y. Each entry in the difference columns is formed by taking the algebraic difference of the two entries on the left. The series in the successive columns are known as the 1st, 2nd, 3rd, 4th, etc., differences respectively.

This table may be continued indefinitely upwards, downwards, and to the right.

In the following illustrations it is assumed that x advances by

unit intervals of the argument i e that $h = 1$ This condition can always be secured by the introduction of an auxiliary variable as shown in Table 59

The differences on the top (sloping) line are known as the Leading Differences Knowing these and the differences in the last column the table can be built up exactly

Let us apply this method to a simple function say $y = x^3$, from $x = -2$ to $x = +6$

TABLE 55
DIFFERENCES OF $y = x^3$

Argument x (1)	Entry y (2)	Differences Δy (3)	$\Delta^2 y$ (4)	$\Delta^3 y$ (5)	$\Delta^4 y$ (6)
2	-8				
		7			
1	1		6		
		1		6	
0	0		0		0
		1		6	
1	1		6		0
		7		6	
2	8		12		0
		19		6	
3	27		18		0
		37		6	
4	64		24		0
		61		6	
5	125		30		
		91			
6	216				

Here $a = -2$ and $h = 1$

$$y_0 = -8 \quad y_1 = -1 \quad y_2 = 0 \text{ etc}$$

$$\Delta y_0 = y_1 - y_0 = -1 - (-8) = 7$$

$$\Delta y_1 = y_2 - y_1 = 0 - (-1) = 1$$

$$\Delta y_2 = y_3 - y_2 = 1 - 0 = 1$$

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = 1 - 7 = -6$$

$$= y_2 - 2y_1 + y_0 = 0 + 2 - 8 = -6$$

$$\Delta^2 y_1 = \Delta y_2 - \Delta y_1 = 1 - 1 = 0$$

$$= y_3 - 2y_2 + y_1 = 1 - 0 - 1 = 0$$

$$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0 = 0 + 6 = 6$$
$$= \Delta y_2 - 2\Delta y_1 + \Delta y_0 = 1 - 2 + 7 = 6$$
$$= y_3 - 3y_2 + 3y_1 - y_0 = 1 - 0 - 3 + 8 = 6$$

$$\Delta^4 y_1 = y_5 - 4y_4 + 6y_3 - 4y_2 + y_1$$
$$= 27 - 32 + 6 + 0 - 1 = 0 \text{ etc., etc.}$$

The laws of formation of Finite Differences follow the Binomial Theorem.

Higher Differences.

In Table 55 the fourth differences disappear exactly. This is because y is a rational integral function of x.

If y is an analytical function of x, not of this description, the higher differences usually tend to disappear.

As an example consider the function $y = 10 \tan x + 5$.

TABLE 56

DIFFERENCES OF $y = 10 \tan x + 5$

Argument	Entry	Differences				
x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Deg. 60	22·321					
		+ 0·719				
61	23·040		+ 0·048			
		+ 0·767		+ 0·004		
62	23·807		+ 0·052		+ 0·002	
		+ 0·819		+ 0·006		− 0·001
63	24·626		+ 0·058		+ 0·001	
		+ 0·877		+ 0·007		0
64	25·503		+ 0·065		+ 0·001	
		+ 0·942		+ 0·008		+ 0·002
65	26·445		+ 0·073		+ 0·003	
		+ 1·015		+ 0·011		
66	27·460		+ 0·084			
		+ 1·099				
67	28·559					

Here $a = 60°$ and $h = 1°$. Each column of differences is smaller than the one preceding, but since $\tan x$ is a transcendental function, there is no complete disappearance.

Referring to Table 55 on page 180 we have

$$y_1 = y_0 + \Delta y_0 = -8 + 7 = -1$$
$$y_2 = y_1 + \Delta y_1 = -1 + 1 = 0$$
$$= y_0 + 2\Delta y_0 + \Delta^2 y_0 = -8 + 14 - 6 = 0$$
$$y_3 = y_2 + \Delta y_2 = 0 + 1 = 1$$
$$= y_1 + 2\Delta y_1 + \Delta^2 y_1 = -1 + 2 + 0 = 1$$
$$= y_0 + 3\Delta y_0 + 3\Delta^2 y_0 + \Delta^3 y_0 = -8 + 21 - 18 + 6 = 1$$
$$y_4 = y_0 + 4\Delta y_0 + 6\Delta^2 y_0 + 4\Delta^3 y_0 + \Delta^4 y_0$$
$$= -8 + 28 - 36 + 24 + 0 = 8$$
$$y_5 = y_0 + 5\Delta y_0 + 10\Delta^2 y_0 + 10\Delta^3 y_0 + 5\Delta^4 y_0 + \Delta^5 y_0$$
$$= -8 + 35 - 60 + 60 + 0 + 0 = 27$$

It nay be shown by induction that these theorems are true for all positive integral values of x.

It is then assumed that they are also true for fractional and negative values.

If we apply this principle to a series whose differences terminate, the result will be exact; if, to a series whose differences do not terminate, the result will only be approximate, although in general we may carry the approximation as far as we please.

Interpolation Formulae.

In practice, we calculate intermediate values by means of an **Interpolation Formula.** A collection of suitable formulae is given below without proof—

TABLE 58

COMMON INTERPOLATION FORMULAE

Name of Formula (1)	When Employed (2)	Formula (3)
Newton's	At beginning of Table	$y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{1 \cdot 2}\Delta^2 y_0 + \frac{x(x-1)(x-2)}{1 \cdot 2 \cdot 3}\Delta^3 y_0 + \frac{x(x-1)(x-2)(x-3)}{1 \cdot 2 \cdot 3 \cdot 4}\Delta^4 y_0 + \ldots$ (4)

TABLE 58—*contd*

COMMON INTERPOLATION FORMULAE

Name of Formula (1)	When Employed (2)	Formula (3)
Newton-Gauss	In middle of Table	$y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{1 \cdot 2}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{1 \cdot 2 \cdot 3}\Delta^3 y_{-1} + \frac{(x+1)x(x-1)(x-2)}{1 \cdot 2 \cdot 3 \cdot 4}\Delta^4 y_{-2} + \ldots$ (5)
Stirlings	In middle of Table	$y_x = y_0 + x\frac{\Delta y_{-1} + \Delta y_0}{2} + \frac{x^2}{2}\Delta^2 y_{-1} + \frac{x(x^2-1^2)}{6}\frac{\Delta^3 y_{-2} + \Delta^3 y_{-1}}{2} + \frac{x^2}{24}(x^2-1^2)\Delta^4 y_{-2} + \ldots$ (6)
Newton-Gauss (backward)	At end of Table	$y_x = y_0 + x\Delta y_{-1} + \frac{(x+1)x}{2}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{6}\Delta^3 y_{-2} + \frac{(x+1)x(x-1)(x-2)}{24}\Delta^4 y_{-2} + \ldots$ (7)

Application to Statistics

In order to apply the method of Interpolation to statistics it is assumed that the formulae may be employed with respect to figures which do not obey any definite mathematical law with approximate results. The degree of approximation obtained will depend upon the original figures: the greater their regularity as shown by the diminution of higher differences the more reliable the result.

Interpolation Applied to Life Annuity

The following figures show the value of a life annuity upon a single life aged 20 at rates of interest varying from $2\frac{1}{2}$ to 5 per cent.

Required to calculate the intermediate values at $2\frac{3}{4}$, $3\frac{3}{4}$, and $4\frac{1}{8}$ per cent.

TABLE 59

CALCULATION OF VALUE OF AN ANNUITY UPON A SINGLE LIFE AGED 20, AT VARIOUS RATES OF INTEREST[1]

Rate of Interest (1)	Auxiliary Variable (x) (2)	Annuity Value (3)	Δ (4)	Δ^2 (5)	Δ^3 (6)	Δ^4 (7)
2·5	0	24·145				
			− 2·102			
3·0	1	22·043		+ ·284		
			− 1·818		− ·047	
3·5	2	20·225		+ ·237		+ ·009
			− 1·581		− ·038	
4·0	3	18·644		+ ·199		+ ·006
			− 1·382		− ·032	
4·5	4	17·262		+ ·167		
			− 1·215			
5·0	5	16·047				

To find the value at $2\frac{3}{4}$ per cent ($x = 0{\cdot}5$), we use Newton's formula (No. 4),

$$y_0 = 24{\cdot}145$$

$$y_{0{\cdot}5} = y_0 + {\cdot}5\Delta y_0 - \frac{{\cdot}5 \times {\cdot}5}{1{\cdot}2}\Delta^2 y_0 + \frac{{\cdot}5 \times {\cdot}5 \times 1{\cdot}5}{1.2.3}\Delta^3 y_0$$

$$- \frac{{\cdot}5 \times {\cdot}5 \times {\cdot}1{\cdot}5 \times 2{\cdot}5}{1.2.3.4}\Delta^4 y_0$$

$$= 24{\cdot}145 - {\cdot}5 \times 2{\cdot}102 - {\cdot}125 \times {\cdot}284$$

$$- {\cdot}0625 \times {\cdot}047 - {\cdot}0391 \times {\cdot}009$$

$$= 23{\cdot}055$$

The calculation of the value at $3\frac{3}{4}$ per cent can be effected by the Newton Gauss formula (No. 5).

To save labour, we adjust column (2), so that it runs -2, -1 $0 + 1 + 2 + 3$.

Writing

$$y_0 = 20{\cdot}225 \text{ and } x = 0{\cdot}5 \text{ we have}$$

[1] Inwood's *Tables of Interest and Mortality*, 1930, p. 272.

$$y_x = y_0 + x\Delta y_0 + \frac{x(x-1)}{2}\Delta^2 y_{-1} + \frac{(x+1)x(x-1)}{6}\Delta^3 y_{-1}$$

$$+ \frac{(x+1)x(x-1)(x-2)}{24}\Delta^4 y_{-2} + \ldots$$

$$y_{0.5} = 20.225 - .5 \times 1.581 - .125 \times .237$$

$$+ .0625 \times .038 + .0234 \times .009$$

$$= 19.407$$

As a check let us calculate by Stirling's formula

$$y_x = y_0 + x\frac{\Delta y_{-1} + \Delta y_0}{2} + \frac{x^2}{2}\Delta^2 y_{-1}$$

$$+ \frac{x(x^2-1^2)}{6}\,\frac{\Delta^3 y_{-2} + \Delta^3 y_{-1}}{2} + \frac{x^2}{24}(x^2-1^2)\Delta^4 y_{-2}$$

$$y_{0.5} = 20.225 - .5 \times 1.6995 + .125 \times .237$$

$$+ .063 \times .0425 - .008 \times .009$$

$$= 19.407 \quad \text{(as before)}$$

To find the value at $4\frac{1}{2}$ per cent we write $y_0 = 18.644$ and $x = 0.25$. By the Newton Gauss formula (No. 5)

$$y_{0.25} = 18.644 - .25 \times 1.382$$

$$- .0938 \times .199 + .0390 \times .032$$

$$+ .0171 \times .006$$

$$= 18.281$$

By Stirling's formula

$$y_{0.25} = 18.644 - .25 \times 1.4815$$

$$+ .03125 \times .199 + .0391 \times .018$$

$$- .002 \times .006$$

$$= 18.281 \quad \text{(as before)}$$

Interpolation in Frequency Distribution.

When it is a question of interpolating in a frequency distribution it is better to work with the cumulative numbers. The following table relates to numbers of estates liable to Estate duty, 1929–30. Required to estimate the numbers between £15,000 and £16,000, and so on up to £20,000.

TABLE 60

NUMBERS OF ESTATES LIABLE TO ESTATE DUTY, ENGLAND, 1929–30[1]

Net Value not exceeding (1)	Auxiliary Variable x (2)	Number of Estates (3)	Δ (4)	Δ^2 (5)	Δ^3 (6)	Δ^4 (7)
£ 5,000	− 2	101,669				
			6,375			
10,000	− 1	108,044		− 3,895		
			2,480		+ 2,751	
15,000	0	110,524		− 1,144		− 2,035
			1,336		+ 716	
20,000	+ 1	111,860		− 428		− 648
			908		+ 68	
25,000	+ 2	112,768		− 360		
			548			
30,000	+ 3	113,316				

Take $y_0 = 110{,}524$. The values required are $y_{0\cdot2}, y_{0\cdot4}, y_{0\cdot6}, y_{0\cdot8}$.

Since the figures are purely statistical, the standard error is large and there is no object in carrying the operation beyond (say) second differences.

Using Stirling's formula (No. 6)

$$y_x = y_0 + x\frac{\Delta y_{-1} + \Delta y_0}{2} + \frac{x^2}{2}\Delta^2 y_{-1}$$

we have

$$y_{0\cdot2} = 110524 + \cdot2 \times 1908 - \cdot02 \times 1144$$

$$= 110883$$

$$y_{0\cdot4} = 110524 + \cdot4 \times 1908 - \cdot08 \times 1144 = 111{,}196$$

[1] *Seventy-third Report of Commissioners of His Majesty's Inland Revenue for the Year Ending 31st March*, 1930 (Cmd. 3802).

The rest of the table can then be built up from the leading differences as follows—

TABLE 61
INTERPOLATED DISTRIBUTION OF ESTATES

Net Value of Estate (1)	Number of Estates (2)	Δ (3)	Δ^2 (4)
£			
15 000	110 524		
		359	
16 000	110 883		−46
		313	
17 000	111 196		46
		267	
18 000	111 463		46
		221	
19 000	111 654		−45
		176	
20 000	111 860		

Interpolation by Unequal Intervals of the Argument

When the argument proceeds by unequal intervals the above formulae are not applicable. In that case the most convenient formula is Lagrange's, viz.—

$$
\begin{aligned}
y_x = y_0 \frac{(x - x_1)(x - x_2) \ldots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)} \\
+ y_1 \frac{(x - x_0)(x - x_2) \ldots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \ldots (x_1 - x_n)} + \ldots \\
+ y_n \frac{(x - x_0)(x - x_1) \ldots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \ldots (x_n - x_{n-1})}
\end{aligned}
\tag{8}
$$

The following Table relates to statistics of super tax payers 1928–9.

TABLE 62
INTERPOLATION BETWEEN SUPER TAX STATISTICS 1928–29

Income not exceeding		Number of Persons	
(1)	x (2)	(3)	y (4)
£			
2 000	x_{-1}	0	y_{-1}
2 500	x_0	23 435	y_0
3 000	x_1	39 818	y_1
4 000	x_2	59 157	y_2
5 000	x_3	69 974	y_3

Required to estimate the number of persons with incomes not exceeding £3,500.

Taking $x_0 = £2{,}500$ and $y_0 = £23{,}485$, etc., etc., we have

$$= 23{,}485 \frac{(3\cdot5-3)(3\cdot5-4)(3\cdot5-5)}{(2\cdot5-3)(2\cdot5-4)(2\cdot5-5)}$$

$$+ 39{,}818 \frac{(3\cdot5-2\cdot5)(3\cdot5-4)(3\cdot5-5)}{(3-2\cdot5)(3-4)(3-5)}$$

$$+ 59{,}157 \frac{(3\cdot5-2\cdot5)(3.5-3)(3\cdot5-5)}{(4-2\cdot5)(4-3)(4-5)}$$

$$+ 69{,}974 \frac{(3\cdot5-2\cdot5)(3\cdot5-3)(3\cdot5-4)}{(5-2\cdot5)(5-3)(5-4)}$$

$$= -4{,}697 + 29{,}863\cdot5 + 29{,}578\cdot5 - 3{,}498\cdot7 = 51{,}246$$

Inverse Interpolation.

In this case it is required to find the value of x corresponding with a given value of y.

Let it be required to find the median of the super-tax distribution in Table 62.

The total number of super-tax payers in 1928–29 was 97,696. It is therefore a question of finding the income of the 48,848th tax-payer.

Let x represent that income, then by Lagrange's formula we have

$$23{,}485 \frac{(x-3)(x-4)(x-5)}{(2\cdot5-3)(2\cdot5-4)(2\cdot5-5)}$$

$$+ 39{,}818 \frac{(x-2\cdot5)(x-4)(x-5)}{(3-2\cdot5)(3-4)(3-5)}$$

$$+ 59{,}157 \frac{(x-2\cdot5)(x-3)(x-5)}{(4-2\cdot5)(4-3)(4-5)}$$

$$+ 69{,}974 \frac{(x-2\cdot5)(x-3)(x-4)}{(5-2\cdot5)(5-3)(5-4)} = 48{,}848$$

This reduces to $x^3 - 14\cdot3039x^2 + 97\cdot4869x - 123\cdot8989 = 0$.

Solving by Horner's method we have $x = 3{,}378\cdot6$.

Therefore the median of the distribution is £3,379.

III Gradlation (or Smoothing)

The subject of Graduation is extremely difficult and treatment in this volume will be confined to a few simple theorems

The typical problem of graduation is to remove accidental fluctua tions from an observed series supposed to follow a regular law Mortality is a function of age and if we tabulate the mortality rates at successive ages on the basis of observed numbers living and dying during a year or period of years the resulting series will show a definite trend marred however by fluctuations of sampling Since it is unthinkable that these fluctuations should be real the series must be smoothed before it can be used for actuarial purposes and it is the object of graduation to remove these disturbances in a systematic manner doing as little violence as possible to the observed facts

The same principle applies to the smoothing of a frequency dis tribution or an historical series

The methods of smoothing by freehand and moving averages have already been discussed [1] The freehand method involves an arbitrary element whilst the moving average misplaces the curve whenever the trend possesses a marked curvature [2]

A Simple Graduation Formula

Generally speaking a smoothing formula replaces the original observation by a graduated observation formed by a combination of the observation itself and those preceding and following

A simple formula is

$$y'_x = y_x - \frac{3}{35}\Delta^4 y_{x-2} \qquad (9)$$

where y_x represents the original and y'_x the graduated value

As an illustration of method let us apply this formula to the following table showing the ratio between legitimate births and relevant marriages 1892 1923 The ratio in question depends upon a complicated series of estimates and it is improbable that the true fig[illegible]ated so much as the estimates indicate

3 000
4 000
5 000

[1] [illegible] 45 and Chapter IX p 68

[2] [illegible] to the base line the moving average locates the [illegible]rve convex to the base line it locates it too high

TABLE 63

RATIO BETWEEN LEGITIMATE BIRTHS AND RELEVANT MARRIAGES, 1892–1923. ORIGINAL ESTIMATES AND GRADUATION PROCESS[1]

Year x (1)	Original Data y_x (2)	Δ (3)	Δ^2 (4)	Δ^3 (5)	Δ^4 (6)	$-\frac{3}{35}\Delta^4$ (7)	Graduated y'_x (9)
1892	4·24	+ ·04					*4·362*
1893	4·28	− ·16	− ·20	+ ·47			*4·262*
1894	4·12	+ ·11	+ ·27	− ·45	− ·92	+ ·079	4·199
1895	4·23	− ·07	− ·18	+ ·22	+ ·67	− ·057	4·173
1896	4·16	− ·03	+ ·04	− ·06	− ·28	+ ·024	4·184
1897	4·13	− ·05	− ·02	+ ·05	+ ·11	− ·009	4·121
1898	4·08	− ·02	+ ·03	− ·11	− ·16	+ ·014	4·094
1899	4·06	− ·10	− ·08	+ ·14	+ ·25	− ·021	4·039
1900	3·96	− ·04	+ ·06	− ·02	− ·16	+ ·014	3·974
1901	3·92	0	+ ·04	− ·08	− ·06	+ ·005	3·925
1902	3·92	− ·04	− ·04	+ ·03	+ ·11	− ·009	3·911
1903	3·88	− ·05	− ·01	− ·05	− ·08	+ ·007	3·887
1904	3·83	− ·11	− ·06	+ ·15	+ ·20	− ·017	3·813
1905	3·72	− ·02	+ ·09	− ·18	− ·33	+ ·028	3·748
1906	3·70	− ·11	− ·09	+ ·24	+ ·42	− ·036	3·664
1907	3·59	+ ·04	+ ·15	− ·32	− ·56	+ ·048	3·638
1908	3·63	− ·13	− ·17	+ ·22	+ ·54	− ·046	3·584
1909	3·50	− ·08	+ ·05	− ·08	− ·30	+ ·026	3·526
1910	3·42	− ·11	− ·03	+ ·08	+ ·16	− ·014	3·406
1911	3·31	− ·06	+ ·05	+ ·02	− ·06	+ ·005	3·315
1912	3·25	+ ·01	+ ·07	− ·14	− ·16	+ ·014	3·264
1913	3·26	− ·06	− ·07	− ·14	0	0	3·260
1914	3·20	− ·27	− ·21	+ ·28	+ ·42	− ·036	3·164
1915	2·93	− ·20	+ ·07	− ·31	− ·59	+ ·051	2·981
1916	2·73	− ·44	− ·24	+ ·64	+ ·95	− ·081	2·649
1917	2·29	− ·04	+ ·40	− ·26	− ·90	+ ·077	2·367
1918	2·25	+ ·10	+ ·14	+ ·62	+ ·88	− ·075	2·175
1919	2·35	+ ·86	+ ·76	− 2·05	− 2·67	+ ·229	2·579
1920	3·21	− ·43	− 1·29	+ 1·47	+ 3·52	− ·302	2·908
1921	2·78	− ·25	+ ·18	− ·02	− 1·49	+ ·128	2·908
1922	2·53	− ·09	+ ·16				*2·800*
1923	2·44						*2·700*

The scheme of calculation is self-explanatory. The four terminal figures are not provided by the formula; they have, therefore, been estimated on the best basis possible.

If the graduated figures are not considered satisfactory, they may be smoothed again by a second application of the formula.

This method removes minor irregularities successfully, but leaves major irregularities untouched. In order to produce a perfectly

[1] Connor, L. R. *Fertility of Marriage and Population Growth*, *J.R.S.S.*, Vol. LXXXIX (1926), p. 562.

smooth curve it would be necessary to employ a more powerful formula such as Spencer's, which is based upon twenty-one terms instead of five For details of this and similar methods the reader is referred to specialized works on the subject

IV Curve Fitting

Another method consists in fitting a mathematical curve by Factorial Moments This is considered in the section following

When the original data run irregularly, it is sometimes desirable to replace them by a continuous curve The usual approach to the subject is by the method of Least Squares Provided however the data run by equal intervals of the argument, it saves time to use Factorial Moments The following is an introduction to the subject

Let there be two variables x and y The variable x is supposed to advance by unit intervals whilst y is supposed to vary in an irregular fashion and it is required to replace the values of y by a regular curve

Consider the following table Column (1) shows the successive values of x and column (2) the corresponding values of y Columns (3) (4) (5) and (6) are formed by successive summation viz $57 + 46 = 103$ $103 + 135 = 295$ etc Each column after the third stops one interval short of the preceding column The values denoted by S_0 S_1 S_2 S_3 are the Factorial Moments and express intrinsic properties of the data

TABLE 64
FITTING A CURVE BY FACTORIAL MOMENTS
The y Data

x (1)	y (2)	Sum of Col (2) (3)	Sum of Col. (3) (4)	Sum of Col (4) (5)	Sum of Col (5) (6)
1	57	57	57	57	57
2	46	103	160	217	274
3	32	135	295	512	786
4	119	254	549	1 061	1 847
5	113	367	916	1 977	3 824
6	115	481	1 398	3 375	7 199
7	107	589	1 987	5 362	12 561
8	207	796	2 783	8 145	20 706
9	232	1 028	3 811	11 956	32 662
10	234	1 262	5 073	17 029	$(= S_3)$
11	224	1 486	6 559	$(= S_2)$	
12	321	1 807	$(= S_1)$		
		$(= S_0)$			

TABLE 65

FITTING A CURVE BY FACTORIAL MOMENTS

The u Data

x (1)	u (2)	Sum of Col. (2) (3)	Sum of Col. (3) (4)	Sum of Col. (4) (5)	Sum of Col. (5) (6)
1	a	a	a	a	a
2	$a+b$	$2a+b$	$3a+b$	$4a+b$	$5a+b$
3	$a+2b+c$	$3a+3b+c$	$6a+4b+c$	$10a+5b+c$	$15a+6b+c$
4	$a+3b+3c+d$	$4a+6b+4c+d$	$10a+10b+5c+d$	$20a+15b+6c+d$	$35a+21b+7c+d$
5	$a+4b+6c+4d$	$5a+10b+10c+5d$	$15a+20b+15c+6d$	$35a+35b+21c+7d$	$70a+56b+28c+8d$
6	$a+5b+10c+10d$	$6a+15b+20c+15d$	$21a+35b+35c+21d$	$56a+70b+56c+28d$	$126a+126b+84c+36d$
7	$a+6b+15c+20d$	$7a+21b+35c+35d$	$28a+56b+56c+28d$	$84a+126b+126c+84d$	$210a+252b+210c+120d$
8	$a+7b+21c+35d$	$8a+28b+56c+70d$	$36a+84b+126c+126d$	$120a+210b+252c+210d$	$330a+462b+462c+330d$
9	$a+8b+28c+56d$	$9a+36b+84c+126d$	$45a+120b+210c+252d$	$165a+330b+462c+462d$	$495a+792b+924c+792d$
10	$a+9b+36c+84d$	$10a+45b+120c+210d$	$55a+165b+330c+462d$	$220a+495b+792c+924d$	$(=S_3)$
11	$a+10b+45c+120d$	$11a+55b+165c+330d$	$66a+220b+495c+792d$	$(=S_2)$	
12	$a+11b+55c+165d$	$12a+66b+220c+495d$	$(=S_1)$		
		$(=S_0)$			

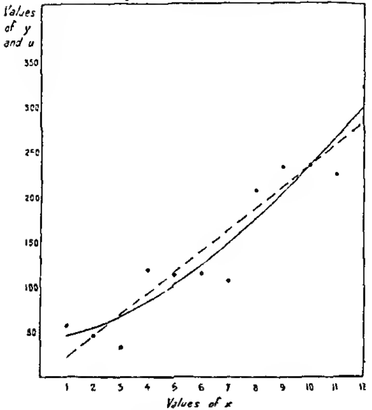


FIG 40

For example, the coefficient of b in S_5 is

$$\binom{n}{5} = \frac{n!}{5!\,(n-5)!} = (\text{for } n = 12)\ \frac{12!}{5!\,7!}$$

$$= \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 792 \quad \text{For } n = 9 \text{ its value is}$$

$$\frac{9!}{5!\,4!} = 126$$

When the value of n is large (say, exceeding 20), the values of the coefficients are also large, and the solution of the simultaneous equations becomes very troublesome. One method of avoiding the difficulty is to average the data in small groups of (say) three items, thus reducing the effective number of observations to one-third. Another method is to fit the curve in sections and join up freehand or by mathematical methods.

The number of coefficients used to determine the values of u is, of course, not limited to four. But since each extra coefficient means an extra summation in order to give a determinate solution of the equations, the method becomes unwieldy when applied to curves higher than the cubic. For most economic data, a straight line fit gives as much as the data will bear, and the elementary student is not recommended to attempt anything more ambitious unless he is reasonably sure there is something to be gained for his trouble.

CHAPTER XVIII

MISCELLANEOUS THEOREMS AND METHODS

I THE PARETO CURVE

If a cumulative frequency distribution of incomes be plotted upon a double logarithmic scale the points will lie approximately upon a straight line[1] This statement is true of *Great Britain* the *United States* *Germany* *British India* and other countries where the law has been tested

The following Table and graph illustrate the application of this proposition to *Great Britain and Northern Ireland*

TABLE 67

GREAT BRITAIN AND NORTHERN IRELAND—CUMULATIVE DISTRIBUTION OF INCOMES 1928–29[2]

Income (x) (1)	Number of Incomes of £x or over (y) (2)	Log (1) (3)	Log (2) (4)
2 000	97 696	3·3010	4·9899
2 500	74 211	3·3979	4·8705
3 000	57 878	3·4771	4·7625
4 000	38 539	3·6021	4·5859
5 000	27 722	3·6990	4·4428
6 000	20 975	3·7782	4·3217
7 000	16 544	3·8451	4·2186
8 000	13 317	3·9031	4·1244
10 000	9 163	4·0000	3·9620
13 000	4 595	4·1761	3·6623
20 000	2 781	4·3010	3·4442
25,000	1 851	4·3979	3·2674
30 000	1 324	4·4771	3·1219
40 000	753	4·6021	2·8768
50 000	487	4·6990	2·6873
73 000	234	4·8751	2·3692
100 000	130	5·0000	2·1139

Column (1) shows the income (x) and column (2) the number of incomes of £x or over (y) Columns (3) and (4) show the respective

[1] This is the so-called Pareto's Law

[2] *Seventy third Report of the Commissioners of Inland Revenue* (1930) Cmd 3802 p 88

logarithms of these quantities. Plotting in logarithms, we arrive at the result shown in Fig. 41.[1] The approximation to a straight line is fairly close.

Algebraic Treatment.

Writing $x =$ income in £

and $y =$ number of incomes of £x or over;

the equation of a straight line is given by

$$\log y = \log N - \alpha \log x \qquad (1a)$$

$$\text{or } y = Nx^{-\alpha} \qquad (1b)$$

The quantity α (Alpha) is known as the slope of the curve: its usual value is about 1·5.

It can be proved that the average of all incomes above x is

$$\frac{\alpha}{\alpha - 1} x \qquad (2)$$

whence it follows that the total income above

$$x = \frac{\alpha}{\alpha - 1} xy = z \text{ (say)} \qquad (3)$$

The total amount of income recorded for the 97,696 persons assessed in 1928–29 was £541,319,350 or £5,541 per head. Therefore the average income above £2,000

$$= \frac{\alpha}{\alpha - 1} x = \frac{\alpha}{\alpha - 1} \times 2000 = £5541$$

whence $\alpha = 1{\cdot}565$

Substituting in equation (1a) we have

$$\log 97{,}696 = \log N - 1{\cdot}565 \log 2000$$

$$\text{or } 4{\cdot}9899 = \log N - 5{\cdot}1661$$

whence $\log N = 10{\cdot}1560$

Therefore, Pareto's equation for the data in question is

$$\log y = 10{\cdot}1560 - 1{\cdot}565 \log x$$

This equation is represented by the dotted line in Fig. 41.

[1] The curve is sometimes plotted the other way round, viz. with income along the y axis and numbers along the x axis. Since, however, we regard x as the independent variable, Fig. 41 shows the correct plot.

Great Britain & Northern Ireland
Distribution of Incomes over £2000
Year 1928 29

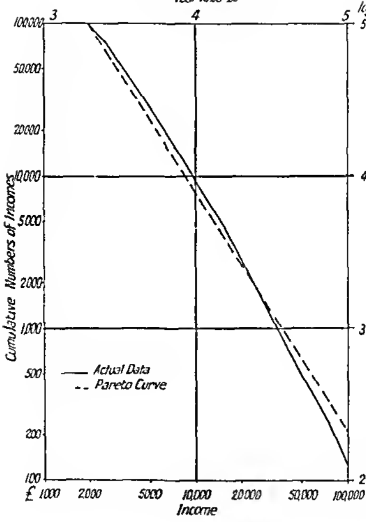


Fig. 41

The steeper the slope of the curve the more equally is income distributed and *vice versa*.

It was once supposed that the PARETO curve represented an ultimate law of income distribution, so that any attempts to alter the existing distribution must necessarily be defeated. It is now held that the curve represents an empirical law, valid within the range of experimental observation, but invalid outside. Therefore, it can be used with confidence for interpolation inside the range but cannot be used for extrapolation.

II. LORENZ CURVE

A variant of the Cumulative Frequency Graph is often employed in order to measure the concentration of wealth or income.

Consider the following data relating to the distribution of estates exceeding £10,000 in net capital value.

TABLE 68

NUMBERS AND CAPITAL VALUES OF ESTATES IN GREAT BRITAIN LIABLE TO ESTATE DUTY, 1929–30[1]

Capital Value exceeding (1)	Cumulative Number of Estates (2)	Col. (2) as Percentage (3)	Cumulative Net Capital Values (4)	Col. (4) as Percentage (5)
(£000)			(£000,000)	
3,000	2	*0·02*	12·4	*3·39*
2,000	6	*0·07*	16·2	*4·42*
1,500	10	*0·11*	24·7	*6·74*
1,000	15	*0·17*	32·7	*8·93*
800	20	*0·23*	36·0	*9·83*
600	35	*0·40*	47·1	*12·86*
500	48	*0·55*	52·6	*14·36*
400	68	*0·78*	60·1	*16·41*
300	119	*1·37*	77·5	*21·16*
250	158	*1·81*	86·4	*23·59*
200	214	*2·46*	100·0	*27·31*
150	317	*3·64*	118·4	*32·33*
100	581	*6·67*	149·5	*40·82*
80	817	*9·38*	169·7	*46·34*
60	1,172	*13·46*	195·2	*53·30*
50	1,467	*16·84*	211·7	*57·81*
40	1,971	*22·63*	233·8	*63·84*
30	2,804	*32·19*	262·3	*71·63*
25	3,420	*39·26*	279·8	*76·41*
20	4,418	*50·72*	302·7	*82·66*
15	5,923	*68·00*	329·6	*90·00*
10	8,710	*100·00*	366·2	*100·00*

[1] *Seventy-third Report of Commissioners of Inland Revenue for the Year ended 31st March*, 1930 (Cmd. 3802), 1931.

Columns (3) and (5) of this Table are plotted in Fig 44 This figure shows that more than 16 per cent of the total wealth passing in 1929–30 was held by less than 1 per cent of decedents and more than 57 per cent by less than 17 per cent of decedents The straight line denotes the line of equal distribution Evidently the concavity of the curve away from the straight line is a measure of concentration of wealth

GREAT BRITAIN DISTRIBUTION OF ESTATES OF DECEDENTS IN 1929 30 LORENZ CURVE

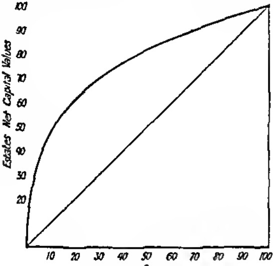


FIG 42

III THE CHI SQUARED TEST

This is a device for testing correspondence between observation and theory

Table 69 is based upon the data of Table 35 (page 123) The method consists in squaring the differences between the actual and theoretical figures dividing by the theoretical figures and summing the results

Small groups have been clubbed together so as to give a minimum of 10 entries

The quantity χ^2 provides a compendious test of the differences

TABLE 69

RIGHT ASCENSION OF POLARIS—ACTUAL AND THEORETICAL FREQUENCIES

Frequency		Column (1) − Column (2) (3)	Column (3) Squared (4)	Column (4) ÷ Column (2) (5)
Actual (1)	Theoretical (2)			
14	14	—	—	—
25	22	+ 3	9	0·4091
43	46	− 3	9	0·1957
74	82	− 8	64	0·7805
126	121	+ 5	25	0·2066
150	152	− 2	4	0·0263
168	163	+ 5	25	0·1534
148	147	+ 1	1	0·0068
129	112	+ 17	289	2·5804
78	72	+ 6	36	0·5000
33	40	− 7	49	1·2250
12	29	− 17	289	9·9655
1,000	1,000			16·0493 $= \chi^2$

between columns (1) and (2). Reference to appropriate tables shows that the chance of the distribution in column (1) having arisen by random sampling from a population of the form of column (2) is about 14 per cent. This chance is not unlikely, and the discrepancy therefore is not regarded as statistically significant.

IV. CONTINGENCY

Consider the following (hypothetical) table supposed to relate to results of differential treatments in some experiment—

TABLE 70

FREQUENCIES OF SUCCESS AND FAILURE FOR VARIETIES OF TREATMENT

Treatment (1)	A (2)	B (3)	C (4)	Total (5)
Successes . .	215	325	60	600
Failures . .	135	175	90	400
TOTAL .	350	500	150	1,000

APPENDIX I

REFERENCES PART I

Aitken A C *On the Graduation of Data by the Orthogonal Polynomials of Least Squares* (Proceedings of Royal Society of Edinburgh 1933)

Barlow s *Tables of Squares Cubes Square Roots Cube Roots and Reciprocals* (Spon London 1930)

Boddington A L *Statistics and Their Application to Commerce* (Pitman London 1941)

Bowley A L *An Elementary Manual of Statistics* (Macdonald & Evans London 1928)

Bowley A L *Elements of Statistics* (King London 1937)

Brunt D *The Combination of Observations* (Cambridge Press 1931)

Davies G R and Crowder W F *Methods of Statistical Analysis in the Social Sciences* (Wiley New York Chapman and Hall London 1933)

Dawson Shepherd *An Introduction to the Computation of Statistics* (University of London Press 1933)

Elderton W P *Frequency Curves and Correlation* (Layton London 1927)

Ezekiel M *Methods of Correlation Analysis* (Wiley New York 1930)

Fisher Irving *The Making of Index Numbers* (H Mifflin Boston 1922)

Fisher R A *Statistical Methods for Research Workers* (Oliver & Boyd Edinburgh 1936)

Hill A Bradford *Principles of Medical Statistics* (*The Lancet* 1937)

Jones D C and Daniels G W *Elements of Mathematics for Students of Economics and Statistics* (Hodder & Stoughton London 1926)

Jones D C *A First Course in Statistics* (Bell London 1929)

Karsten K G *Charts and Graphs* (Prentice Hall New York 1923)

King W I *Elements of Statistical Method* (Macmillan New York 1920)

Levy H and Roth L *Elements of Probability* (Oxford Press 1936)

Macaulay F G *The Smoothing of Time Series* (National Bureau of Economic Research New York 1931)

Mellor *Higher Mathematics for Students of Chemistry* (Longmans 1922)

Mills F C *Statistical Methods* (Pitman London 1929)

Milne Thompson L M *The Calculus of Finite Differences* (Macmillan London 1933)

Milne Thomson L M and Comrie L J *Standard Four figure Mathematical Tables* (Macmillan London 1931)

Pearl Raymond *Introduction to Medical Biometry and Statistics* (Saunders Philadelphia 1930)

Pearson E S *The Application of Statistical Methods to Industrial Standardisation and Quality Control* (British Standards Institution London 1936)

Pearson K *Tables for Statisticians and Biometricians* (Cambridge Press Part I 1914 Part II 1931)

Pride James *Seven figure Mathematical Tables* (Chambers 1929)

Rhodes E C *Elementary Statistical Methods* (Routledge London 1933)

Rietz H L (Editor) *Handbook of Mathematical Statistics* (Mifflin Boston 1924)

Secrist H *An Introduction to Statistical Methods* (Macmillan New York 1925)

Shewhart W A *Economic Control of Quality of Manufactured Product* (Van Nostrand New York 1931)

Tippett L H C *The Methods of Statistics* (Williams & Norgate London 1937)

Whittaker E T and Robinson G *The Calculus of Observations—A Treatise on Numerical Mathematics* (Blackie London 1932)

Woods H M & Russell W T *An Introduction to Medical Statistics* (King London 1931)

Yule G U and Kendall M G *An Introduction to the Theory of Statistics* (Griffin London 1937)

CHAPTER XIX

INTRODUCTION

The first part of this work has dealt with elementary Statistical Methods, including simple tests of the reliability of results. This course gives the reader access to a large range of non-technical and semi-technical statistical literature, and affords a useful introduction to works of an advanced character, intended for professional statisticians.

The second part introduces the reader to the application of these methods under actual working conditions. For a variety of reasons, published statistics, i.e. data published by official bodies and private bodies enjoying a recognized status, are to be preferred to unpublished statistics at this stage.

(1) Published statistics relate to matters of public interest and so possess greater generality.

(2) The original sources together with details as to compilation are open to all, so that readers are immediately in a position to verify any matters on which they are in doubt.

(3) The space gained in the present work by referring the reader for details to the original sources instead of printing the data in full becomes available for comment and explanation.

In view of students' requirements, the illustrations will be drawn for the most part from economic and administrative data, with especial reference to current problems.

For treatment of more specialized problems, the reader is referred to such works as Carr-Saunders and Jones's *Survey of the Social Structure of England and Wales*, Fisher's *Statistical Methods for Research Workers*, and Newsholme's *Vital Statistics*.

The reader will notice that most of these illustrations are based upon straightforward arithmetic and that comparatively little use is made of more complicated notions, such as dispersion and the theory of correlation. The reason is that economic and administrative data do not usually belong to the class known as "statistically uniform" and they will not therefore bear the strain of refined analysis.

Supplementary Reading

Students should supplement this course by their own reading The most useful current publications are as follows—

The Bank of England Statistical Summary
The Board of Trade Journal
The Ministry of Labour Gazette
The Statistical Abstract for the United Kingdom
The Abstract of Labour Statistics
The Economist—Trade Supplement (Monthly)
The Statist
The Journal of the Royal Statistical Society
The Economic Journal
The Bulletins of the London and Cambridge Economic Service
The Publications of the League of Nations and the International Labour Office
Trends (now incorporated with *Industry Illustrated*)

Official Statistics

By far the greater part of current statistical information is derived from official sources[1] Government departments possess far greater facilities for collecting reliable information than private individuals and they alone are usually able to bear the expense of publication

Official Statistics Must be Used with Care and Caution Statistics collected *ad hoc* are usually organized with a view to comparability Much information is however in the nature of by products of the administrative machine and is not suitable for scientific purposes without inquiry and adaptation

A Central Statistical Bureau

Most foreign countries possess Central Statistical Bureaux charged with collecting and editing all statistical matter of public interest A proposal to set up such a bureau for the United Kingdom has been discussed but has been dismissed as impracticable Every department is responsible for its own statistics but co-ordination is effected by a *Permanent Consultative Committee on Official Statistics* who are responsible for the preparation of the *Guide to Current Official Statistics* The *Board of Trade* also exercise co-ordinating functions by virtue of their responsibility for the issue of the *Statistical Abstract*

[1] Consult Bowley *Official Statistics and How to Use Them* the *Guide to Current Official Statistics* and the *Report of the Committee on Industry and Trade—Further Factors in Industrial and Commercial Efficiency* p 268 *et seq*

Non-official Statistics.

Many trade associations collect statistical data of production, sales, prices, etc., from their members, but the returns are regarded as highly confidential, and the public have no access to them. So much importance is attached to secrecy that arrangements are frequently made for the collection of returns under code numbers only known to the organization's executive. Frequently the tabulation is done by an outside firm of accountants who do not know the names of the firms and the totals only are passed to the secretary of the association, who has no access to the details. By such elaborate means is perfect secrecy preserved.

Attention should, however, be drawn to the figures of output of pig-iron, steel, and rolling mill products published monthly by the *National Federation of Iron and Steel Manufacturers*; figures relating to output and work in hand of shipyards published by *Lloyd's Register*; figures as to the progress of the cotton trade published by the *International Federation of Master Cotton Spinners' Associations*; and the figures of activity in the Electrical Industry published by the *British Electrical and Allied Manufacturers' Association.*

Statistics and Standardized Costing Systems.

The *Committee on Industry and Trade* have commented on the difficulty of obtaining costing data in comparable form on account of the lack of consistent and scientific practice among many firms in respect of cost accounts. A striking feature of the reports of the Committees under the Profiteering Acts was the frequency of reference to the unsatisfactory organization of businesses in this respect.

Although the position is improving, it is by no means satisfactory, even in some of the largest and most important industries, and there is little uniformity of principle in systems throughout a given industry.

A few trade associations are seeking to introduce a greater measure of uniformity among their members. One important association has its own standardized method of costing, devised to suit the peculiarities of the industry, and its members are advised as to the adaptation of this system to their individual works. Members make periodical returns of their costs, and also half-yearly returns showing their oncost rates constructed on the lines approved by the association. The effect of fluctuations in the oncost rates on the costs of various articles is gone into; average

rates constructed from these returns are sent out to members as well in order that the individual efficiency may be compared with the average

Another association has engaged an expert with a view to setting up a standard of costing practice and an examination is made of the systems employed by firms producing a variety of products

The Committee report that the variety of conditions precludes any attempt to lay down a uniform costing system applicable to all classes of business enterprise but that an important work lies before the trade associations and pioneer firms in the various industries in persuading the more backward firms to realize the importance of scientific costings without which the most efficient and economic management is impossible

Addendum (1947) Attention is drawn to the *Monthly Digest of Statistics* the new *Annual Abstract of Statistics* and the annual *White Paper on the National Income and Expenditure of the United Kingdom* all produced by the Central Statistical Office and on sale at H M Stationery Office

CHAPTER XX

POPULATION[1]

THE census is a count or stocktaking of the entire population at midnight on an appointed date, *usually in the spring, when fewest* people are away from their homes.

This count is taken every ten years. Statutory authority exists for a quinquennial census, but the Government has not availed itself of its powers under the Act.

The officer responsible for the census and for the systematic registration of all births, deaths and marriages in England and Wales is the Registrar-General, who works under the general direction of the Minister of Health. The Registrar-General's headquarters and permanent staff are located in London at Somerset House. For both census and registration purposes the country is divided into registration districts and subdistricts, each in charge of a local official. For census purposes a further subdivision into enumeration districts and an additional staff of some 40,000 enumerators are required. As their work lasts only a few weeks, temporary employment only is afforded to these enumerators. On engagement, shortly before the appointed date, the enumerator's first duty is to make a thorough examination of his district. He then sets out to deliver schedules, compiling, as he goes, an exhaustive record of every building or set of premises and every household residing in them. He visits every household and leaves a schedule to be filled in by the householder, who is liable to a penalty of £10 if he refuses to make a return or gives false information.

The householder's schedule is a form on which he must fill in statutory particulars with respect to every person in the dwelling alive at midnight on the appointed date. The particulars required in 1931 were as follows—

(*a*) Name and surname of every person alive at midnight of Sunday, 26th April.

[1] This chapter relates to the Census of England and Wales. Similar arrangements are in force for the rest of the United Kingdom and for the British Dominions.

(*b*) Relationship to head of household
(*c*) Usual residence
(*d*) Sex
(*e*) Age
(*f*) Condition as to marriage
(*g*) Birthplace
(*h*) Nationality
(*k*) Personal occupation
(*l*) Employer and employer s business
(*m*) Particulars of persons over 14 not following an occupation for profit

To be filled in by the enumerator—

No of rooms
No of males
No of females
No of persons

These particulars always differ between one census and another since apart from staple inquiries there has always been a fringe of additional inquiries admitted as of importance on particular occasions A separate form is required for every family A boarder visitor or servant is counted in with the family but a lodger who boards separately is counted as a separate family Shortly after the appointed day the enumerator calls again gives the householder any necessary assistance to complete the form himself adds up the number of persons of either sex and determines and fills in the number of rooms When he has completed his quota of schedules he puts them together makes out a summary and sends them in to the local census officer The schedules for the whole country are sent to London where their contents are codified transferred to punched cards analysed and mechanically tabulated Preliminary results appear a few weeks after the census date and final results are published gradually (1) in a series of county volumes and (2) in a series of summary volumes for the whole country Tabulations must be elaborate in order to utilize all this information effectively and the whole operation takes several years to complete

Special care has to be exercised as to the description of occupations and industries and special codes of instructions are issued to enumerators with a view to securing uniformity

Contents of the Census Volumes.

The contents of the Census Volumes are as follows—

1. Preliminary Report, including tables of the population enumerated in England and Wales (Administrative and Parliamentary Areas) and in Scotland, the Isle of Man and the Channel Islands.
2. County Volumes—Part I. Areas as constituted at the date of the Census. (Each volume contains introductory letterpress and tables forming a general statistical survey of the county concerned.)
3. County Volumes—Part II. Areas as constituted under Orders giving effect to the review of County Districts under the Local Government Act, 1929.
4. Ecclesiastical Areas (England).
5. Occupation Tables.
6. Classification of Occupations.
7. Industry Tables.
8. Classification of Industries.
9. Housing.
10. General Tables, including Population, Institutions, Ages and Marital Conditions, Birthplace and Nationality, Welsh Language.

The following summary tables are published in the *Statistical Abstract for the United Kingdom* (Comd. 5353) and the *Abstract of Labour Statistics for the United Kingdom* (Comd. 5556).

In the *Statistical Abstract for the United Kingdom*—

1. Population of each division of the United Kingdom at each Census from 1821.
2. Estimated population at the middle of each year, 1913 and 1922–36.
3. Births, deaths and national increase, 1870–1935.
4. Age, sex, marital distribution, and birthplaces at each Census, 1871–1931.
5. Houses occupied and unoccupied at each Census 1851–1931.
6. Numbers of persons enumerated in occupations at Censuses of 1921 and 1931 in Great Britain and of 1926 in Northern Ireland.
7. Numbers and proportions in the principal industry groups at each Census from 1881 in Great Britain.

In the *Abstract of Labour Statistics for the United Kingdom*—

8. Persons gainfully occupied; numbers enumerated and percentages in various age groups in Great Britain in 1911, 1921, and 1931.
9. Numbers of persons aged 14 and over enumerated in certain groups of occupations in Great Britain at the Census of 1931, classified as (*a*) managerial, (*b*) operative, (*c*) working on own account, and (*d*) out of work.
10. Numbers and proportions, aged 10 years and upwards, engaged in the principal industry groups in Great Britain at successive Censuses, 1891–1931, and intercensal percentage changes.
11. Numbers, aged 14 years and upwards, enumerated in certain industries in Great Britain in 1931.

Owing to differences in content and classification the Census Statistics according to industry do not, generally speaking, agree with those published by the *Ministry of Labour*.

Other Publications of the General Register Office.

The most important of these are—

1. Weekly Returns of Births and Deaths in England and Wales
2. Quarterly Returns of Births and Deaths in England and Wales
3. The Registrar-General's Annual Statistical Review of England and Wales.

For detailed explanations of the statistics referred to in this chapter, the reader is referred to (*a*) the original publications and (*b*) the standard works on vital statistics. (See list of References.)

CHAPTER XXI

PRICES

STATISTICS of prices are plentiful, and details may be found in the Technical, Trade, and Financial Journals. For studies of prices in the past, reference should be made to the standard works on the subject.[1]

Any systematic and extensive study of prices necessitates the use of **Price Index Numbers**, and it is the object of this Chapter to describe and illustrate the principal indices employed in this country.

The Board of Trade Index Number of Wholesale Prices.[2]

The present series of index numbers begins with the month of January, 1935, and displaces an older series dating from 1920. The construction of an index of wholesale prices involves the selection of a representative series of commodities and the assignment to these of weights; in other words, the relative importance of each commodity represented in the index has to be settled according to some plan. As wide a range of commodities as can be secured is desirable, but in practice the choice is limited to those for which price data are regularly available. These comprise in the main raw materials and semi-manufactured goods, fully finished articles, except for certain foodstuffs, not being sold on exchanges. An index number of wholesale prices cannot therefore be representative of goods in all stages of manufacture. Wholesale prices of fully finished goods tend to move in the same manner, though not to the same extent, as the prices of the principal materials entering into their composition, and in assigning weights to those materials, account is taken of the added importance which they derive as constituents of the finished goods. Accordingly the index may show the direction of the movement in prices of all

[1] E.g. Layton and Crowther's *Introduction to the Study of Prices*.

[2] The following account is based upon articles in the *Board of Trade Journal* for 20th January, 1921 (p. 61); 24th April, 1930 (p. 551); 10th December, 1931 (p. 739); and 24th January, 1935 (Supplement). See also Flux, "The Measurement of Price Changes," *J.R.S.S.*, Vol. LXXXIV (1921), p. 167; and "The Measurement of Price Changes: Retrospect and Prospect," *J.R.S.S.*, Vol. XCVI (1933), p. 606 *et seq.*

goods sold at wholesale but not the extent, and a direct comparison cannot be made between the movements of an index number of wholesale prices and one of retail prices

Method of Construction

The total number of commodities included is 200, and the total number of quotations 258 the difference being due to the fact that in some cases the average of two or more quotations is used in order to obtain a more representative figure The commodities include foodstuffs and materials of industry and semi manufactured products and are arranged in eleven groups In averaging the percentage price changes from the base year for the several commodities the geometric means of those changes not their arithmetic means have been used The geometric mean has the effect of reducing the influence of upward movements in price and increasing that of downward movements modifications that correspond to the decreased or increased consumption that is likely to accompany such price movements The principal advantage of the geometric mean is however that it enables a change of base year to be made without affecting the proportionate change in the general index or in any group index It secures that an index is reversible, so that the relative price change between 1920 and 1930 for example is the same irrespective of whether 1920 or 1930 is used as base year The method has the further advantage that it enables quotations which have ceased to be representative of the course of prices to be replaced by more representative quotations without affecting the balance of the index

The data obtained at the 1930 Census of Production have been used in the compilation of the index The basis adopted has been to make the weights[1] used as nearly as possible proportional to the total value of goods manufactured or produced within the United Kingdom together with the imports of goods of the same description which pass into consumption without undergoing a process of manufacture after importation Duplication between the various trades comprised within a group has been eliminated as has also the duplication between groups resulting from the inclusion

[1] The index is not weighted in the strict and formal sense but inside the groups an equivalent effect is obtained by increasing the number of quotations in respect of commodities of special importance The eleven groups are considered to be of equal importance and are averaged on that footing

in one group of a commodity which clearly forms a dominant material in another group. The quotations used are for the most part weekly (in some cases daily) quotations taken from published journals, but in a few cases the *Board of Trade* are supplied with wholesale prices by important firms engaged in the manufacture of commodities for which published particulars are not regularly available. The prices of dutiable goods are inclusive of the appropriate amount of the duty. The weekly quotations are combined so as to obtain monthly averages, and for each month the average price change compared with that of the corresponding month of the preceding year is computed. Thus a continuous series of index numbers is obtained. In calculating mean annual figures, the geometric average of the index numbers for the twelve months is taken.

The base year has been successively 1913, 1924, and 1930. It may be expected that in the near future the base will be changed to 1935, the year of the last Census of Production.

The Eleven Groups.

The contents of the eleven groups are as follows. (The figures in parenthesis show the number of quotations used.)

I. Cereals (20). II. Meat, Fish and Eggs (20). III. Other Food and Tobacco (28). IV. Coal (9). V. Iron and Steel (37). VI. Non-ferrous Metals (8). VII. Cotton (10). VIII. Wool (11). IX. Other Textiles (9). X. Chemicals and Oils (15). XI. Other Articles (33). Totals: Food and Tobacco (68), Industrial (132). Grand total (200).

In addition to the index numbers for the various groups combined into a total for all articles, index numbers for a grouping of the 119 industrial items other than fuel into *basic materials, intermediate products*, and *manufactured articles* are prepared. The index for manufactured articles should not be regarded as an index for finished goods, since most of these are for commodities which are subjected to further manufacturing processes before entering into consumption. The absence of quotations for fully finished goods results from the practical impossibility of securing quotations for finished articles which are precisely comparable in quality and character over a period of years. A further group index, relating to building materials, is compiled from the various quotations for materials used in the group index.

Specimen Tables.

The results are published monthly in the *Board of Trade Journal* in the form shown in Tables 72 and 73

From the above account it will be seen that the price relatives

TABLE 72

BOARD OF TRADE INDEX OF WHOLESALE PRICES FOR THE MONTH OF MAY 1937 WITH COMPARATIVE FIGURES[1]

Group	No of Items	Increase (+) or Decrease (−) per Cent in May 1937 compared with		Index Numbers (1930 = 100)		
		April 1937	May 1936	May 1937	April 1937	May 1936
(1)	(2)	(3)	(4)	(5)	(6)	(7)
I Cereals	20	− 2.5	+ 39.1	126.3	129.5	90.8
II Meat fish and eggs	20	+ 2.9	+ 10.3	86.4	84.0	78.3
III Other food and tobacco	28	− 0.8	+ 4.1	95.0	95.8	94.1
Total—Food and tobacco	68	0.3	+ 15.3	101.7	102.0	88.2
IV Coal	9	− 0.8	+ 19.2	125.0	126.0	104.9
V Iron and steel	37	+ 11.1	+ 25.5	131.4	118.3	104.7
VI Non ferrous metals	8	− 4.9	+ 36.8	123.4	129.8	90.2
VII Cotton	10	− 2.2	+ 24.8	106.7	109.1	85.5
VIII Wool	11	+ 0.3	+ 37.5	136.3	136.1	103.0
IX Other textiles	9	+ 0.3	+ 10.7	78.3	78.1	70.7
X Chemicals and oils	15	0.6	+ 9.0	100.1	100.7	91.8
XI Miscellaneous	33	+ 0.6	+ 25.2	113.8	113.1	90.9
Total—Industrial materials and manufactures	132	+ 2.6	+ 23.0	115.4	112.5	93.8
Total—All articles	200	+ 1.7	+ 20.5	110.7	108.9	91.9
Industrial materials (excluding fuel)—						
Basic materials	33	0.5	+ 36.9	131.6	132.2	96.1
Intermediate products	37	+ 3.2	+ 21.7	110.9	107.5	91.1
Manufactured articles	49	+ 4.8	+ 17.2	113.6	108.4	96.9
Building materials	16	+ 0.5	+ 9.5	104.0	103.5	95.0

[1] *B T J* 10th June 1937 p 805

TABLE 73

BOARD OF TRADE INDEX OF WHOLESALE PRICES—MAY, 1936–MAY, 1937[1]

Averages for the Year 1930 = 100

Group	1936								1937				
	May	June	July	Aug.	Sept	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
I Cereals	90·8	87·3	92·0	105·6	110·5	114·3	112·0	118·1	123·1	121·6	124·1	129·5	126·3
II. Meat, fish and eggs	78·3	80·9	81·2	82·5	82·6	84·0	82·5	85·0	82·2	81·9	83·0	84·0	86·4
III. Other food and tobacco	94·1	97·4	95·2	94·2	94·4	98·0	97·5	98·1	97·9	96·9	99·6	98·8	98·0
Total—Food and tobacco	88·2	89·3	89·9	93·7	95·0	98·0	96·7	99·3	99·4	98·6	100·7	102·0	101·7
IV. Coal	104·9	105·0	105·0	104·9	106·6	107·5	108·2	111·2	112·7	117·7	123·0	126·0	125·0
V. Iron and steel	104·7	106·4	108·1	108·4	108·7	108·8	108·9	110·5	112·0	112·6	115·4	118·3	131·4
VI. Non-ferrous metals	90·2	88·4	88·9	90·2	92·0	94·7	102·0	106·6	113·2	121·1	142·9	129·8	123·4
VII. Cotton	85·5	87·2	92·4	90·5	99·3	92·2	94·0	94·5	97·3	99·6	107·0	109·1	106·7
VIII. Wool	103·0	101·3	101·1	103·0	103·7	104·8	114·2	122·4	129·4	127·3	129·9	136·1	136·5
IX. Other textiles	70·7	70·4	70·7	72·5	72·2	73·3	74·3	75·2	75·9	75·9	76·9	78·1	78·3
X. Chemicals and oils	91·8	92·3	93·2	93·2	93·7	93·7	94·6	97·1	99·5	99·5	100·5	100·7	100·1
XI. Miscellaneous	90·9	91·5	92·3	92·7	93·8	94·5	95·3	97·4	101·8	105·9	110·1	113·1	113·8
Total—Industrial materials and manufactures	93·8	94·3	95·5	96·0	96·6	97·3	99·1	101·5	104·6	106·5	110·7	112·5	115·4
Total—All articles	91·9	92·6	93·6	95·2	96·1	97·6	98·3	100·8	102·9	103·9	107·3	108·9	110·7
Industrial materials (excluding fuel)													
Basic materials	96·1	95·9	97·4	98·6	99·1	100·5	104·9	109·8	116·7	120·9	129·2	132·2	131·6
Intermediate products	91·1	91·5	93·5	93·9	95·0	95·8	97·2	100·4	103·3	104·3	107·0	107·5	110·9
Manufactured articles	96·9	98·1	98·8	98·9	99·2	99·6	100·4	100·9	102·2	103·2	106·5	108·4	113·6
Building materials	95·0	96·1	96·9	97·2	97·9	98·5	99·3	100·3	101·1	101·2	103·8	103·5	104·0

[1] *B.T.J.*, 10th April, 1937, p. 804.

are calculated upon the chain base[1] principle each figure being calculated upon the corresponding figure of the previous year In this form they are adapted to current commercial requirements For comparative purposes the figures are chained up and expressed as a percentage of the average for the year 1930 The use of the chain base avoids the difficulties which occur when owing to changes in business commodities once serving as standards of comparison are superseded by other commodities or other grades The extension of the list of commodities when necessary is facilitated in the same way, the calculations from any date not being hampered by the necessity of securing comparisons with prices at a past date from which the calculations have started The results are not therefore dependent upon the choice of base year In general the results of proceeding from one year to another directly and by means of the chain will be identical but this statement needs qualification in the event of the introduction of additional (or alternative) price relatives

The "Economist" Index Number of Wholesale Prices

This index was begun in 1864 and has been twice revised in 1911 and 1928 In its present form the index is based upon the unweighted geometric mean of changes in the prices of 58 commodities (the base year being 1927) Subsidiary indices are calculated for Cereals and Meat Other Foods Textiles Minerals and Miscellaneous

The results are published as follows—

(1) *Monthly*—Complete index and subsidiaries (with date base 1929) in the *Monthly Trade Supplement*

(2) *Fortnightly*

(*a*) Complete index and subsidiaries (date base 1927)

(*b*) Complete index only (date base 1913 and 1924)

(*c*) Complete index compared with indices of *prices of primary products* (British sterling and American dollar) price of gold (sterling) and wholesale prices in four other countries with date base 18th September 1931 (the date of leaving the Gold Standard)

The fortnightly figures are published in the main paper under the heading *Notes of the Week*

[1] See Chapter XVI p 160

The "Statist" Index Number of Wholesale Prices.

The *Statist* Index forms a continuation of a series begun by the late Mr. Augustus Sauerbeck. The base data are averages for the period 1867–77, and the number of commodities is forty-five. These are arranged in six groups, as shown in the following specimen calculation for 1930—

TABLE 74

CALCULATION OF *Statist* INDEX NUMBER FOR 1930

	Index Numbers	1867–77 Total Numbers	Example for 1930	
			Total Numbers	Average
(1)	(2)	(3)	(4)	(5)
1. Vegetable food, corn, etc. (wheat, flour, barley, oats, maize, potatoes, and rice)	8	800	619	77
2 Animal food (beef, mutton, pork, bacon, and butter) . . .	7	700	992	142
3 Sugar, coffee, and tea . . .	4	400	215	54
(1)–(3) *Food* . .	19	1,900	1,826	96
4. Minerals (iron, copper, tin, lead, and coal)	7	700	784	112
5. Textiles (cotton, flax, hemp, jute, wool, and silk)	8	800	669	84
6. Sundry materials (hides, leather, tallow, oils, soda, nitrate, indigo, and timber)	11	1,100	1,071	97
(4)–(6) *Materials* . .	26	2,600	2,524	97
General Average . .	45	4,500	4,350	97

The *Statist* Index Number has not undergone any substantial reconstruction, and it is now presented in substantially the same form as it originated. In opposition to the usual practice, its compilers publish annually full details of its construction. These two features make the *Statist* Index peculiarly valuable for experimental purposes where a continuous record of figures over a long period is required.[1]

[1] Full details are published annually in the *J.R.S.S.* The above description is based upon the account appearing in *J R.S.S.*, Vol. XCIV, 1931, p. 267; and Vol C, 1937, p. 277.

TABLE 25

Statist INDEX NUMBER OF WHOLESALE PRICES BY GROUPS 1924–36
(Base = Average 1867–77)

Year	Vegetable Food (Corn etc.)	Animal Food (Meat etc.)	Sugar Coffee and Tea	Total Food	Minerals	Textiles	Sundry Materials	Total Materials	Grand Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1924	119	153	105	130	158	170	120	146	139
1925	118	162	89	128	154	165	119	143	136
1926	108	150	88	119	151	133	114	131	126
1927	108	138	83	114	141	131	118	129	122
1928	107	142	78	114	133	136	117	124	120
1929	99	146	72	110	126	122	111	119	115
1930	77	142	54	96	113	84	97	97	97
1931	68	119	50	83	100	63	85	82	83
1932	72	105	50	[illegible]	99	64	81	81	80
1933	60	106	47	74	107	67	80	83	79
1934	63	108	50	77	109	72	80	85	82
1935	66	107	42	[illegible]	112	80	93	90	84
1936	[illegible]	109	41	81	118	83	90	94	88

Other Wholesale Price Indices

These include—

(1) The *Financial Times* index published weekly and monthly

(2) Reuter's Index quoted in the *Financial News* This is a daily index of commodity prices and represents an attempt to measure the world price level

(3) *The Times* index published monthly

Ministry of Labour Cost of Living Index Number[1]

This figure is designed to measure the average increase in the cost of maintaining unchanged the pre war standard of living of the working classes i e the standard actually prevailing just before the War irrespective of whether or not such standard was adequate

Pre war expenditure of course varied widely in different cases according to total family income and it is therefore convenient to express the increase in the cost of living in the form of a percentage

which can be applied to different amounts of pre-war expenditure. Even where the total weekly expenditure was the same in different families, however, the distribution of this expenditure over different commodities varied considerably according to the number of persons in the family, their ages, sex, and mode of living; and as some articles have risen less in price than others, even the percentage increase would vary to some extent with different families. It is obvious, therefore, that no single figure can apply exactly to individual cases, and that if the increase is to be expressed in the form of a single percentage, the only practicable method is to calculate a fair general average.

Owing to variations in the amounts of increase in the prices of different commodities, economies or readjustments in expenditure have no doubt been effected in many families, especially in those cases where incomes have not increased so much as prices; on the other hand, the standard of living has probably been raised in many families in which wages have been increased in greater proportion than prices.

No account is taken in the figures of any such alterations in the standard of living, as to which trustworthy statistics are not available.

Items Included.

The items included in the statistics fall into five main groups, viz. food, rent, clothing, fuel and light, and miscellaneous items.

FOOD. The foodstuffs included are beef, mutton, bacon, fish, flour, bread, potatoes, tea, sugar, milk, butter, margarine, cheese, and eggs. Fruit and vegetables (other than potatoes) are omitted on account of the difficulty of obtaining continuous and comparable quotations.

Information is collected at the beginning of each month by the managers of employment exchanges and branches from representative retailers (including co-operative societies, large multiple firms, and private shopkeepers) doing a working-class trade. The total number of retailers is over 5,000, distributed among over 500 towns and villages.

From this information is calculated the average percentage increase in price over July, 1914, for each article. The percentages so obtained are then weighted by figures based upon the average

TABLE 76

MINISTRY OF LABOUR—CALCULATION OF AVERAGE PERCENTAGE INCREASE IN COST OF FOOD OVER JULY 1914 AS AT 1ST JANUARY 1931

Article (1)	Average Expenditure on these Articles in Budgets of 1904 (2)	Weights proportional to such Expenditure (3)	Percentage Increase at 1st January 1931 (4)	Result of Multiplying (3) by (4) (5)
	s d			
Beef	2 5¾	48	39	1 872
Mutton	1 2¾	24	51	1 224
Bacon	11¾	19	17	323
Fish	3½	9	112	1 008
Flour	1 0¾	20	21	420
Bread	2 6¾	50	29	1 450
Tea	1 1½	22	28	616
Sugar	11¾	19	21	399
Milk	1 5¼	25	84	2 100
Butter	2 1½	41	16	656
Cheese	6½	10	38	380
Margarine	—	10	3	30
Eggs	1 –	19	96	1 824
Potatoes	11	18	28	504
	16 7¾	334	—	12 806

Weighted average increase = 12 806 ÷ 334
= 38 per cent

expenditure shown by 1 944 working-class family budgets collected by the *Board of Trade* in 1904 [1]

Table 76 illustrates the compilation of the figure representing increase in cost of food as at 1st January 1931

RENT Information is obtained periodically from town clerks property owners associations and house agents in a number of large towns respecting increases in rents in controlled and decontrolled dwellings and the ratios in which these two classes of dwellings stand From these figures a final figure is compiled representing the average percentage increase in rent for the whole country

CLOTHING Information is obtained as to retail prices of men s suits and overcoats woollen and cotton materials underclothing

[1] Inquiries made from time to time indicate that there was little change between 1904 and 1914 in the proportion of income spent on different commodities and that *therefore the original weights are valid for a 1914 base* The only exceptional item is margarine for which a special allowance has been made

and hosiery, and boots, as generally bought by the working classes before the War, i.e. relatively low-priced grades. Inquiry forms distributed and collected through the post are completed each month by 300 representative outfitters, drapers, and boot retailers in eighty-one towns. The descriptions of articles for which quotations are given vary with different retailers, but before the form is dispatched to a retailer the prices quoted by him at the previous inquiry are entered on it, and he is asked to quote the current prices for the same articles and qualities as before. The price relatives are worked up on the chain base principle,[1] and are then combined into a weighted average in which allowance is included for the cost of "making up" garments, in cases in which materials are purchased.

FUEL AND LIGHT. Returns are obtained from coal merchants, gas undertakings, and retailers in a number of towns, and the percentage increases shown are combined into a weighted average upon principles similar to those indicated above, except that a fixed base is employed instead of a chain base.

OTHER ITEMS. These include soap and soda, domestic ironmongery, brushware and pottery, tobacco and cigarettes, fares and newspapers. Information is obtained from retailers, transport undertakings, and (as regards newspapers) the Press. The various percentages of increase shown are combined into a weighted average on the fixed base principle.

Combination of Results.

The figures for the five groups of items are combined into a general weighted average, the weights being food, $7\frac{1}{2}$; rent (including rates), 2; clothing, $1\frac{1}{2}$; fuel and light, 1; and other items $\frac{1}{2}$. These weights represent proportionate expenditures as ascertained from the budgets referred to above, the results of an inquiry into rents in 1912, and other available information.

Table 77, shown on page 226, shows the compilation of the final figure as at 1st December, 1930.

The cost of living figure is presented in the form of an average percentage increase in the cost of living, not as an index number. In order to convert it into the latter, it is necessary to add 100, e.g. an average increase of 55 per cent in cost of living corresponds with an index number of 155.

[1] See Chapter XVI, p. 160

TABLE 77

UNITED KINGDOM—AVERAGE INCREASE AS COMPARED WITH JULY 1914 IN WORKING-CLASS COST OF LIVING

Item	Weight	Percentage Increase 1st December 1930	Col (2) × Col (3)
(1)	(2)	(3)	(4)
Food	7·5	41	307·5
Rent (including rates)	2·0	54	108·0
Clothing	1·5	105	157·5
Fuel and Light	1·0	75	75·0
Other items	0·5	75	37·5
	12·5	—	685·5

Average percentage increase = 685·5 ÷ 12·5 = 55 per cent (say)

Items Not Included in the Statistics

The list of items included is considered sufficiently extensive and representative to provide a sound basis for estimating the average increase in cost of living for a working-class family. Items not included form only a small proportion of total working class weekly expenditure and their omission would influence only the final result if the price of the omitted items (taken together) were either very much below or very much above the general average. So far as can be judged it is unlikely that the general average increase would be appreciably affected by the inclusion of a larger number of items even if it were found practicable to extend the list.

principle of marginal utility, i.e. that she will spend less money upon articles that become relatively dear and more upon articles that are relatively cheap, whilst maintaining the total amount of satisfaction. If this proposition be accepted, then the current method of compilation overstates the percentage increase. Others question whether a cost of living index ought to reflect such changes, and argue that a working-class family cannot, in fact, make them without loss of nutrition or satisfaction.

6. Change in constitution of the family. Families are smaller than formerly, and less expenditure is incurred in rearing children. On the other hand, there is a prospect of reduction in supplementary earnings.

7. Failure to allow for benefits of social services, which have increased substantially since 1914.

In answer to a question in the House of Commons on 7th April, 1936, asking the Minister of Labour whether he proposed to revise the basis of the cost of living index number, the Minister said—

> Yes, Sir. I have recently given further consideration to this matter, and have decided that a revision of the basis of the cost of living index number should now be undertaken. For this purpose it will be necessary to collect data with regard to the distribution of the main items of expenditure of working-class households at the present time. An inquiry of this character, on a scale sufficiently comprehensive to provide representative information covering different seasons of the year, cannot be completed before the end of next year. In the meantime the cost of living index number will continue to be calculated on the existing basis, and I anticipate that the new index number can be so linked on to the previous numbers as to continue the series without a break. I should add that, as regards the methods to be adopted in the conduct of the inquiry, I hope to have the assistance of a small advisory committee, which will include representatives of employers and trade unions.

The Minister also stated the terms of reference of the Advisory Committee, which are as follow—

> To advise the Minister of Labour as to the methods to be adopted in the collection of information, by means of family budgets, showing the approximate average weekly expenditure of working-class families on the items which should be taken into account in the construction of index numbers, designed to measure the percentage changes, from month to month, in the cost of maintaining a present-day standard of living.

It is proposed that about 30,000 working-class households, distributed over all parts of Great Britain, shall be visited and

that these are unrepresented the picture is defective. Nevertheless, it is not entirely inadequate.

Weighting of Index Numbers.

This raises the awkward question of the weights to be applied. If the number of items included were very large, then weights might possibly be dispensed with. In the absence of this *desideratum*, weighting is desirable, but it is not clear whether it should be based upon quantities produced or quantities consumed or quantities that change hands. Moreover, some commodities are used in the manufacture of other commodities, and it is not clear whether and how, adjustments should be made on this account.

As a practical expedient it is usual to weight by quantities produced, for statistics of consumption and exchange are indefinite or unobtainable. Again lack of information has forced the compilers of index numbers to rely upon base year weighting, in spite of the theoretical advantages obtainable by crossed indices or hybrid weights.[1] Much research is needed and much improvement must be effected before the method of Index Numbers has been placed upon a really sound basis.

Comparisons—United Kingdom

Table 78 gives a comparison between the movements of the *Board of Trade*, the *Economist* and the *Statist* Indices of Wholesale Prices for the period 1920 to 1936. The three wholesale price figures differ in the amount of movement registered but generally speaking agree as to the times in which changes of direction occurred. The differences widen in periods of violent price movement and close up in periods of comparative stability.

The *Cost of Living Index* follows a somewhat different course. Retail prices undergo less violent movements than wholesale prices and there is a time lag. Distribution costs enter largely into the retail price figure, and it is supposed that these have been kept up by the maintenance of wages in sheltered industries at an uneconomic level.

The *Board of Trade* figures are based upon average prices for each month of the year, whereas the *Economist* and the *Statist* figures are based upon average figures for the end of each month and the

[1] See Chapter XVI p 170

Cost of Living figures upon the averages for the beginning of each month. In a detailed and accurate study of the figures it would be necessary to adjust for these discrepancies. As, however, Table 78 is presented only for illustrative purposes, it is considered advisable to ignore these complications.

TABLE 78

WHOLESALE PRICES AND COST OF LIVING, 1924–36

Year	Wholesale Prices					
	Board of Trade		*Economist*		*Statist* (Average 1867–77 = 100)	Cost of Living *Ministry of Labour* (July, 1914 = 100)
	(1913 = 100)	(1930 = 100)	(1913 = 100)	(1927 = 100)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			A—Indices as published			
1924	166·2	—	159·3	—	139	175
1925	159·1	—	154·2	—	136	176
1926	148·1	—	143·2	—	126	172
1927	141·6	—	137·6	100	122	167½
1928	140·3	—	135·1	98·1	120	166
1929	136·5	—	127·2	92·4	115	164
1930	119·5	100·0	106·8	77·6	97	158
1931	104·2	87·8	89·3	64·9	83	147½
1932	101·6	85·6	86·1	62·6	80	144
1933	100·9	85·7	86·8	63·1	79	140
1934	104·1	88·1	90·3	65·6	82	141
1935	—	89·0	94·3	68·6	84	143
1936	—	94·4	100·1	72·8	88	147
			B—Converted to a common date base (1930 = 100)			
1924	—	139·1	—	149·3	143	111
1925	—	133·1	—	144·4	140	111
1926	—	123·9	—	134·1	130	109
1927	—	118·5	—	128·8	126	106
1928	—	117·4	—	126·5	124	105
1929	—	114·2	—	119·1	119	104
1930	—	100·0	—	100·0	100	100
1931	—	87·8	—	83·6	86	93
1932	—	85·6	—	80·6	82	91
1933	—	85·7	—	81·3	81	89
1934	—	88·1	—	84·6	85	89
1935	—	89·0	—	88·2	87	91
1936	—	94·4	—	93·7	91	93

International Price Comparisons—Wholesale Prices

International comparisons of wholesale price movements are given in the *League of Nations* Monthly Bulletin of Statistics and Statistical Year Books

At the time of writing (July 1937) the tables show 46 wholesale price series for 43 countries

The published results show considerable variations in base year Some of the more important index numbers have therefore been standardized on a common base year and in this connection allowances have been made for currency re valuations

Comparisons between wholesale prices of different countries are difficult to interpret The contents of the indices differ there are the effects of tariffs and foreign trade restrictions to be considered and it is by no means clear how changes in currency standards should be regarded the more especially as gaps frequently arise between a currency s internal and external purchasing power It is evident that the table can be used as a basis for only very broad comparisons

International Comparisons—Cost of Living

The corresponding tables for cost of living show 51 series for 45 countries converted to the same base year (1929) The composition of these series is by no means uniform the number including the full quota of items commonly associated with a cost of living index (foodstuffs heating and lighting clothing rent and miscell aneous) being 38 Allowances have been made as before for cur rency re valuations

These figures are even more difficult to interpret than wholesale prices Standards of living differ enormously all over the world and as the prices concerned are mainly internal prices the steadying influence exercised by international commodities is lacking

An inquiry[1] into relative costs of living in Detroit and 14 European cities has been made by the International Labour Office at the request of the Ford Motor Company Limited (London) with the object of finding the cost of securing to workers in European towns the equivalent standard of living of the workpeople in the Ford undertaking in America

[1] *An International Inquiry into Costs of Living* (King London 1932)

CHAPTER XXII

WAGES

STATISTICS of wages are usually available in the following forms—

1. Wage rates—
   (a) Standard rates.
   (b) Minimum rates.
2. Actual earnings.

Wage rates express the amount normally payable for a specified quantum of work, and the figures should be supplemented, if possible, by details of normal working hours, payments for overtime, bonuses, etc.

Statistics of wage rates may be presented either—

1. In the form of actual figures for each occupation or grade concerned.

2. In the form of index numbers, which may embody adjustments for changes in normal hours of labour and other factors influencing the worker's normal earnings.

Statistics of earnings show the actual amounts earned per week irrespective of the basis of payment. The figures may be expressed in the form of a frequency distribution, and the mean and standard deviation calculated. A commoner plan is to show the median and quartiles.

Detailed studies of wages are occasionally made for particular industries and groups of industries. No outstanding statistical problems are involved in such studies, and the reader is referred to the specialized literature on the subject.

Wage Problems.

Studies affecting the wage level and the wage bill of the country afford matter of greater statistical interest. The two major problems concerned are—

1. Wages considered as costs to the employer.
2. Wages considered as income to the wage earners.

In the former case emphasis is laid upon the productivity side, and both labour in general and particular grades of labour are regarded

merely as factors of production amenable to the principle of substitution Such studies naturally lead to comparisons between wage costs and other costs of production

In the latter case emphasis is laid upon wages as a means of livelihood The wage earner is studied as citizen and consumer and comparisons are instituted between wages and costs of living Unfortunately the material at present under our command does not enable us to effect a satisfactory separation between these two problems and topical discussions in which the distinction is ignored have added to the prevalent confusion

In order to study these two problems effectively it would be necessary to compile a series of national cost accounts with subdivisions for each industry The nucleus of such a system already exists in the shape of periodical Censuses of Production [1] but since employers are not at present compelled to disclose details of their wage bills and other costs a thoroughly satisfactory statistical investigation of the problem is out of the question for the time being

Sources of Information

There are no satisfactory records [2] of the total number of wage earners actually in employment and whilst average earnings in most of the principal industries are known with fair precision information on losses of income due to holidays sickness unemployment and trade disputes is defective Even were these particulars forthcoming we should still be at loss for want of satisfactory information as to other elements of cost and profits

A general view of wage movements may be obtained by the method of index numbers The construction of such an index number will be dealt with later [3] In the meantime it is proposed to give an account of current methods of wage payment and wage adjustment followed by a review of the principal sources of information upon wage questions

[1] See Chapter XXVII

[2] The Population Census is made at decennial intervals the statements made by the informants are not always reliable and do not sufficiently distinguish the sick unemployed and retired nor salaried persons from wage earners The Census of Production figures of employees refer only to productive industry whilst the National Insurance figures exclude some wage-earners and include some salaried persons

[3] See p 241

Systems of Wage Payment—Time Rates.

Time wages are expressed as so much per hour, week, or other period. Nominally, payment does not depend upon work done, but there is often an understanding as to the amount of work to be done within the period, and extra rates are often paid for work involving especial skill, unusual risk, or discomfort. In large organizations it is sometimes the practice to grade the employees with special reference to length of service, e.g. in the police and railway services.

Payment by Results.

The commonest method is individual piece-work, i.e. payment for defined tasks. In some trades the arrangements between employers and employed take the form of elaborate price lists, containing innumerable variations on account of the nature of the material, product, or process, or particulars of the machinery used. Price lists are usually arranged with regard to the amount of work a normal man can do during the week, and there is sometimes a provision that some standard day wage shall be guaranteed to the piece-worker irrespective of output.

In some cases where groups of men are working together, systems of group piece-work are in operation: either the chief member of the group is a sub-contractor who pays his assistants, or the total sum is divided between the different grades in the group in proportions, determined by custom or agreement.

There is also a variety of cases in which an element of time wage is combined with an element of payment by results.

There is, in fact, no defined line of demarcation between the two systems, for time rates often imply a definite amount of work, and piece rates are often arranged to produce a definite aggregate of earnings.

No recent information is available as to the proportions of employees paid on the two systems.

Comments on the Above.

Time rates take the form of definite sums of money, but there is no uniformity as regards the unit of time. Employees at hourly rates will lose in the event of a shortening of working hours unless rates are increased to a compensating extent. Account must also be

The numbers of workpeople affected by these changes are known approximately, and the aggregate results of the changes are tabulated year by year in the following form—

TABLE 80

UNITED KINGDOM CHANGES IN RATES OF WAGES IN INDUSTRY GROUPS 1936[1]

Industry Group	Approximate Number of Separate Individuals Reported as Affected by		Estimated Net Weekly Amount of Change in Rates of Wages		Estimated Net Weekly Increase in Rates of Wages of All Workpeople Affected
	Net Increases	Net Decreases	Increases	Decreases	
(1)	(2)	(3)	(4)	(5)	(6)
			£	£	£
Coal Mining	767 700	—	167 925	—	167 925
Other Mining and Quarrying	34 750	50	3 925	5	3 920
Brick Pottery Glass, Chemical etc	190 750	50	13 800	10	13 790
Iron and Steel	144 250	—	21 000	—	21 000
Engineering	493 500	—	42 450	—	42 450
Shipbuilding	86 900	—	9 850	—	9 850
Other Metal	153 600	—	15 325	—	15 325
Textile	562 250	250	67 125	100	67 025
Clothing	74 200	—	8 550	—	8 550
Food Drink and Tobacco	13 400	—	2 600	—	2 600
Woodworking Furniture etc	67 700	—	9 250	—	9 250
Paper Printing etc	7 100	—	550	—	550
Building Public Works Contracting etc	575 750	—	50 750	—	50 750
Gas Water and Electricity Supply	128 900	50	12 850	10	12 840
Transport	630 500	300	53 100	40	53 060
Public Administration Services	94 250	100	9 275	10	9 265
Other	36 900	—	4 750	—	4 750
Total	4 062 400	800	493 075	175	492 900

In addition to the numbers shown in the above table about 21 000 workpeople received increases and sustained decreases of equal amounts during the year

It should be observed that as the changes in the wage of adult men are usually greater than those affecting women boys and girls comparisons of the average amount of change per head to be deduced from the figures are affected by the varying proportions

[1] *Ministry of Labour Gazette* April 1937 p 130

of men, women and young persons employed in the different groups of industries.

Comparison with Previous Years.[1]

In Table 78 the number of workpeople recorded as affected by changes in rates of wages, and the net amount of increase or decrease in 1936, in the industries for which statistics are available, are shown in comparison with similar figures for previous years. The figures quoted in the table must be regarded in the light of certain qualifications. In the first place, it should be noted that the changes in wage rates reported to the Ministry are in the main those arranged between organized groups of employers and workpeople, and that many changes among unorganized workers, especially those affecting only employees of single firms, are not reported. Moreover, as already stated, certain large groups of workpeople are definitely excluded from the scope of the statistics. In consequence of these limitations the figures should not be regarded as affording more than a general indication of the direction of the movement of wages in any year, and a very rough measure of the extent of such movement in comparison with that of other years; and significance should not be attached to small variations in the amount of change between different years. Further, the fact that the changes reported relate mainly to organized workers results in the figures being influenced, over a series of years, by fluctuations in the strength of the workers' organizations. This is particularly the case during the period since 1914, in which such fluctuations have been very considerable. The movement towards the negotiation of wage changes on a national basis since the war period has also tended to make the figures more comprehensive, for such changes do not escape notice, whereas, when separate arrangements are made in each locality, it is possible that some of the changes, especially among those affecting only the smaller districts, may not be reported. It should also be observed that, during the war period, the number of female workers in industry was above the normal and the number of male workers considerably below normal; and as the amounts of increases or decreases in the rates of wages of female workers are generally smaller than those agreed upon for

[1] The following account is based upon an article in the *Ministry of Labour Gazette*, April, 1937, pp. 131–2.

males in the same industry the aggregate amount of the changes in those years was lower than it would have been if the pre-war proportions of male and female employees had been maintained. The relative levels of wages at the end of 1914 and 1936 therefore cannot be accurately ascertained by deducting the aggregate amount of the reductions from the aggregate amount of increases recorded. The figures, however, afford an indication of the general trend of money rates of wages during the period covered.

TABLE 81

UNITED KINGDOM.—CHANGES IN RATES OF WAGES, 1915–1936

Year	Approximate Number of Separate Individuals Reported as Affected by		Estimated Net Weekly Amount of Change in Rates of Wages		Estimated Net Weekly Increase (+) or Decrease (–) in Rates of Wages of all Workpeople Affected
	Net Increases	Net Decreases	Increases	Decreases	
(1)	(2)	(3)	(4)	(5)	(6)
			£	£	£
1915	4 305 000	—	867 100	—	+ 867 100
1916	4 848 000	250	885 250	50	+ 885 200
1917	6 362 000	75	2 986 200	5	+ 2 986 195
1918	6 924 000	—	3 434 500	—	+ 3 434 500
1919	6 240 000	500	2 547 200	60	+ 2 547 140
1920	7 867 000	500	4 793 200	180	+ 4 795 020
1921	78 000	7 244 000	13 600	6 074 600	6 061 000
1922	73 700	7 653 000	11 450	4 221 500	4 210 050
1923	1 202 000	3 079 000	169 000	486 000	317 000
1924	3 0[illegible]9 000	481 500	616 000	61 100	+ 553 300
1925	873 000	851 000	80 900	159 000	78 100
1926	420 000	740 000	133 000	83 700	+ 49 300
1927	93 000	1 855 000	30 700	388 500	357 800
1928	217 000	1 615 000	21 800	163 800	142 000
1929	142 000	917 000	12 900	91 700	78 800
1930	768 000	1 100 000	59 500	116 100	56 600
1931	47 000	3 010 000	5 150	406 300	401 150
1932	33 500	1 949 000	2 600	251 800	249 200
1933	179 500	894 000	17 250	82 500	65 250
1934	1 344 000	85 500	95 500	4 000	+ 91 500
1935	354 500	49 600	19 500	6 800	+ 189 700
1936	4 0[illegible]2 400	800	493 075	175	+ 492 900

It will be seen that the number of workpeople affected by increases in wage rates in 1936 was the largest since 1920, and the aggregate net weekly increase in rates of wages the largest since 1924.

Wage Rates—United Kingdom.

Information as to wage rates is fairly comprehensive. The *Ministry of Labour* publish elaborate tables showing the relative level of rates of wages for adult workers in the principal industries and occupations at July, 1914, and the end of each year from 1919 onwards.[1]

Inquiries into Earnings.[2]

In October, 1935, an inquiry was instituted by the *Ministry of Labour* into the average weekly earnings and weekly hours of labour of workpeople employed in manufacturing industries generally, and in some of the principal non-manufacturing industries, in Great Britain and Northern Ireland. Inquiries on broadly similar lines, but less detailed in some respects, had previously been undertaken in 1924, 1928, and 1931.

Inquiry forms were addressed to all employers (in the industries covered) employing more than 10 workpeople, and to about 20 per cent of the smaller firms, taken at random, asking for particulars of (1) the total number of wage-earners at work in the week ended 12th October, 1935, distinguishing, so far as possible, the numbers of men (21 years and over), youths and boys, women (18 years and over) and girls; (2) the total wages paid to these workpeople in that week, showing separately, so far as possible, the wages paid to men, youths and boys, women and girls, respectively; (3) the hours of labour in a *full* ordinary week, exclusive of meal times and overtime; (4) the number of workpeople who, in the specified week, were working hours less than the full ordinary week, and the average number of hours lost per head by these workpeople in that week; (5) the number of workpeople who, in that week, were working hours in excess of the full week, and the average number of hours worked by these workpeople, during that week, in excess of the full ordinary week.

Employers were asked to include in their returns the whole of the wage-earners (other than those working at home on material

[1] See Statistical Abstract for the United Kingdom (Cmd. 5353) and the forthcoming Twenty-first Abstract of Labour Statistics, due for publication this year (1937).

[2] The following section is based on an account appearing in the *Ministry of Labour Gazette*, February, 1937, pp. 46–9. Further tables are published in this and subsequent issues of the *Gazette*.

TABLE 82

AVERAGE EARNINGS IN THE MANUFACTURING, ETC., INDUSTRIES—UNITED KINGDOM—WEEK ENDED 12TH OCTOBER, 1935

Industry (1)	Total Number of Workpeople Covered by Returns Received (2)	Average Weekly Earnings of all Workpeople Covered (3)	Men (21 Years and over) Number Covered by Returns (4)	Men (21 Years and over) Average Weekly Earnings (5)	Youths and Boys (Under 21 Years) Number Covered by Returns (6)	Youths and Boys (Under 21 Years) Average Weekly Earnings (7)	Women (18 Years and over) Number Covered by Returns (8)	Women (18 Years and over) Average Weekly Earnings (9)	Girls (Under 18 Years) Number Covered by Returns (10)	Girls (Under 18 Years) Average Weekly Earnings (11)
		s. d.		s. d.		s. d.		s. d.		s. d.
TEXTILE INDUSTRIES										
Cotton spinning and doubling (including thread manufacture)	148 110	32 1	27 2 4	49 6	6 685	18 7	18 141	27 6	10 104	16 7
Cotton weaving	121 938	35	15 366	49 10	2 707	19 9	50 738	30 8	3 839	16 7
Cotton spinning and weaving (not separately distinguished)	38 002	33 0	7 277	49 4	2 144	20 3	14 640	28 1	2 577	15 8
Cotton wool, surgical dressings, ... manufacture	6 820	31 11	983	50 1	257	23 6	1 985	28 8	215	17 5
Total Cotton	322 879	33 8	50 741	49 9	8 735	19 1	95 600	28 8	17 745	16 9
Wool sorting, carbonising and scouring	1 430	30 7	598	56 6	96	21 6	95	26 12	19	10 10
Wool combing and top making	13 302	44 5	6 815	53 6	514	27 9	3 128	28 1	245	22 1
Worsted spinning and weaving	79 166	35 7	11 610	55 8	4 431	21 6	31 509	30 6	6 853	20 4
Woollen spinning and weaving	69 177	41 0	1 216	55 0	2 079	22 2	17 675	33 8	3 091	29 8
Woollen and worsted (not separately distinguished)	15 10	35 11	5 897	57 11	1 204	25 0	8 511	31 0	1 583	19 0
Mungo, shoddy and flock manufacture, rag grinding and carbonising	5,489	39 5	1 522	51 7	166	24 8	970	25 0	280	15 9
Total Woollen and Worsted	191,265	38 8	44 316	55 3	9,149	22 5	81 932	31 3	12,075	19 1
Artificial silk spinning	20 741	48 3	8 159	67 3	2 264	39 11	4 910	29 9	1 591	17 10
Silk throwing, spinning and weaving (including artificial silk weaving)	43 889	42 2	25 158	66 10	2 781	30 8	14,058	31 2	4 468	17 1
Flax and hemp spinning and weaving	30 612	26 0	7 631	45 0	2 295	16 1	20 779	25 6	6,041	14 11
Jute spinning and weaving	20 900	34 10	5,421	45 10	2 980	23 10	9 289	32 3	1 715	17 11
Asbestos manufacture	6,040	47 2	3,010	61 3	2 216	27 9	2 100	35 2	264	18 3
Hair curling, spinning and weaving	1 595	35 10	474	53 4	155	24 4	465	25 7	175	15 11
Preparing, spinning and weaving of other or mixed fibres	1 50	30 3	540	49 7	202	22 2	217	27 9	81	15 4
Hosiery manufacture	55 930	37 10	10 068	72 5	2 545	26 1	33 709	35 6	12 834	17 6
Lace manufacture	7 973	43 3	3 102	64 11	470	21 11	2 998	31 9	578	15 11
Carpet and rug manufacture	22 243	37 11	4 336	62 10	1 244	20 8	5 302	36 9	2 911	17 10
Rope, cord and twine manufacture	12,077	30 2	1 741	53 8	785	18 1	5 559	27 3	1,616	15 11
Tapes and smallwares manufacture	6 942	29 11	628	61 0	201	17 3	2 465	29 10	911	15 2
Elastic web manufacture	2 197	32 7	530	50 8	255	20 3	1 661	30 3	587	15 8
Canvas goods (tents, tarpaulins, etc.) manufacture	6 735	32 0	2 585	56 0	331	22 6	2 619	26 10	2,412	16 1
Hemming and embroidery	5 190	20 9	258	58 6	96	21 5	5 131	28 6	1,083	11 10
Making of other textile goods (not dress)	5 223	28 8	541	56 9	84	17 5	2 937	31 0	1 148	15 0
Textile bleaching, printing, dyeing and finishing	78,859	49 6	43 859	55 6	7 702	24 2	12 310	27 10	3 756	16 10
Velvet and fustian cutting	380	39 5	30	59 0	—	—	180	27 0	—	—
Making up and packing	6 022	33 3	2 206	54 2	325	18 5	2 212	27 1	616	15 4
Miscellaneous textile (including combinations of above)	12 366	32 9	2 977	55 8	594	21 9	4 014	29 9	1 665	16 9
Total Textiles	923 674	36 4	206,383	55 11	44 022	23 1	290,985	32 3	78,354	17 2

supplied by the employer), but to exclude managers, clerks, typists, commercial travellers, shop assistants and salaried persons generally. Foremen, carters, warehousemen, etc., were to be included in the returns. In cases where employment in the week ended 12th October was affected by holidays, breakdown, fire, strike or lock-out, or other exceptional circumstances, employers were asked to substitute particulars for the nearest week of an ordinary character.

The number of establishments to which inquiry forms were issued was about 126,000. About 9,000 of these were found to employ no wage-earners within the scope of the inquiry, and 3,000 supplied returns which were unsuitable for tabulation. The number of effective returns received was approximately 76,000. In view of the voluntary character of the inquiry this response is highly satisfactory, thanks to the employers who furnished information, and to the National Confederation of Employers' Organizations and their affiliated organizations, who co-operated with the Department in the arrangements for making the inquiry. While the proportions of workpeople covered by the inquiry vary in different industries, the returns received are, in general, amply representative to provide a trustworthy indication of the average earnings and hours of labour in the principal industries at the date to which the inquiry related.

Total Wage Bill—United Kingdom.

The total wage bill for the United Kingdom has been estimated by Colin Clark, using the Population Census, 1931, the Census of Production, 1930, and a variety of information from other sources. The calculations are highly complicated and we shall only quote results.

These figures include all manual workers engaged in industry, commerce, transport and public services as well as private domestic servants, but exclude clerical and commercial workers employed at weekly rates, who for this purpose are reckoned as "salaried employees."

Bowley and Stamp's estimate for 1924 (quoted in previous editions of this book) was £1,731,000,000. The difference between the two figures is mainly due to the inclusion of clerks and shop assistants in the latter.

relative proportions of skilled and unskilled workers nor of time and piece workers. The results are published for quarterly periods since 1924 and at intervals for particular dates.[1]

Bowley's Wage Index.

This index was introduced in 1929, and supersedes an earlier index based upon unweighted averages.[2]

A variety of purposes is served by measurement of average wages, and to each purpose corresponds an appropriate definition and method of computation. The most general measurement, that used in estimates of national income, is of the average annual earnings of all manual labourers in the United Kingdom. It is convenient to measure separately lost time due to illness, complete unemployment, and holidays. There remain at least six factors which affect earnings, which need to be treated differently for different purposes, viz.—

(*a*) Changes in time rates.
(*b*) Changes in piece rates.
(*c*) Facilities for earning on piece rates.
(*d*) Weekly hours of work.
(*e*) Shifting of relative numbers and re-grading within an industry.
(*f*) Shifting of relative numbers between industries.

There is also to be considered the labour cost of a given unit of production to the employer, as contrasted with the reward for a normal week's labour to the operative.

The Ministry of Labour's method in constructing its general index number[3] has been to depend upon (*a*) and (*b*) only, ignoring all other factors, but paying some attention to (*d*), and it is the object of the new index to bring the remaining factors into the calculation, so far as this is possible.

The results of the Censuses of Production, 1907 and 1924, are not inconsistent with the proposition that output per head was nearly the same in 1924 as before the War in spite of the reduction of working hours, and that money labour-cost per unit output had

[1] See *Twenty-first Abstract of Labour Statistics* (Cmd 4625) and *Ministry of Labour Gazette*

[2] The following account of Professor Bowley's wage index is based upon *Royal Economic Society Memorandum No. 12, A New Index-number of Wages*.

[3] See p 214

more than doubled since 1907 but rather less than doubled since 1914

The basis of the new index is that the number 195 represents the level in December 1924 (100 in July 1914) This is the best estimate we can make for the change in average earnings of all manual workers in the United Kingdom for the normal week It also represents the change in the money-cost of labour per unit product since the reorganization of industry and the change (if any) in labour effort per hour have apparently compensated for the reduction of hours

The list of data included is as follows—

Bricklayers Average of summer weekly time rates in certain towns

Bricklayers Labourers Average of summer weekly time rates in certain towns

Compositors Weighted averages of weekly time rates

Docks Average of half day s time rates at certain ports

Fitters and Engineering Labourers Time rates in principal centres

Coal Weighted average changes in piece rates

Railways Estimated effect of changes (if any on total wage bill)

Cotton Changes in piece rates

Wool Combination of changes in time and piece rates

Agriculture Average of time wages

Shipbuilding Changes in piece rates

Local Authorities Non-trading Services Average of labourers time rates in certain large towns

Tram Drivers and Conductors Average of time rates in certain large towns

Lorry Drivers Average of time rates in certain large towns

Boot and Shoe making Agreed national minimum time rates for women

Sugar and Confectionery Industries Trade Board general minimum time rates for women

Ready made Tailoring Trade Board general minimum time rates for women

Shirt Making Trade Board general minimum time rates for women

Tobacco. Trade Board general minimum time rates for women.

The changes in each of these twenty groups are expressed as percentages of the levels at the end of 1924, and a weighted average is taken. The weights chosen are as follows.

Bricklayers	11
Bricklayers' labourers	4
Compositors	3
Dock labourers	3
Fitters	12
Engineering labourers	7
Shipbuilding	4
Railways	11
Cotton	10
Wool	5
Local authorities	4
Trams	3
Lorry drivers	3
Boots and shoes	1
Confectionery	1
Tailoring	2
Shirt-making	1
Tobacco	1
Agriculture	4
Coal	10
	100

The ideal weights would be proportional to the weekly wage bill in each industry in the country in 1924. These have been roughly estimated for each of the twenty industries or occupations included, subject to certain adjustments.

The form of index number adopted by Professor Bowley is an average of ratios[1] with base year weights, i.e.

$$P_{01} = \frac{\Sigma(q_0p_0)\dfrac{p_1}{p_0}}{\Sigma(q_0p_0)}.$$

where the p's stand for representative time or piece-rates and the q's are chosen so that quantities like q_0p_0 are proportional to the estimated total wage bill in the industry concerned.[2]

What Bowley's Index Measures.

The next thing to consider is what precisely does this index number measure, and Professor Bowley's comments on this point are illuminating and instructive.

[1] See Chapter XVI, p. 170.

[2] For a full discussion, see Bowley, A. L., "Notes on Index Numbers," *Economic Journal*, Vol. xxxviii, June, 1928, pp. 235–7.

For the period 1914 to 1924 the index reflects a combination of all the factors (a) to (f) mentioned on page 251 From 1924 onwards the index must continue to ignore factors other than (a) and (b) since current information (except for miners) relates only to wage rates They can be trued up from time to time with regard to (d) changes in hours of work and (f) shifting of relative numbers between industries But factors (c) changes in facilities of earning and (e) movements within an industry can be brought in only when from time to time complete estimates of earnings in one or all industries become available

Thus the index measures at December 1924 the change in the average level of a week s *earnings* of all manual workers not wholly unemployed and afterwards it reflects the changes in *rates* primarily till it can be rectified to continue to measure earnings

The whole discussion throws an interesting light upon the capacity of the method of index numbers to solve highly intricate problems

The true data exist but they are confidential and there are no means of enforcing disclosure Even were the information obtainable it would not bear the cost of collection and tabulation

By the method of index numbers it is possible to piece together fragmentary information (provided it is representative) and to arrive at a result sufficiently reliable for practical purposes whilst employing only a small fraction of the staff and incurring only a small fraction of the expense involved by investigations upon a comprehensive scale

Ramsbottom s Wage Index

A new index based on continuous records of changes in wage rates in 63 industries has been compiled by Mr E C Ramsbottom of the *Ministry of Labour* with a view to eventual substitution for the existing index [1]

Real Wages

Division of the money wage index by cost of living gives an index of Real Wages which may be accepted subject to qualifications due to lack of precision in the original series as a measure of the command of the working classes over the necessaries and amenities of Life Table 84 shows the result of dividing Bowley s

[1] *J R S S* Vol xcviii (1935) pp 639–73

wage index (base = December, 1924) by the *Ministry of Labour's* cost of living index (base = July, 1914, converted to December, 1924).[1]

TABLE 84

INDEX OF REAL WAGES—UNITED KINGDOM (1920–36)

December (1)	Wage Index (December, 1924) = 100) (2)	Cost of Living (Converted to December, 1924) = 100 (3)	Index of Real Wages = Col. (2) ÷ Col. (3) (4)
1920	154	148½	103½
1921	124½	110	113
1922	99	99½	99½
1923	96½	98	98½
1924	100	100	100
1925	100½	98	102½
1926	101	99	102
1927	100½	93½	107½
1928	99½	93	107
1929	99	92½	107
1930	98½	85½	115
1931	96½	82	117½
1932	94½	79	119½
1933	94	79	119
1934	94½	79½	118½
1935	95¾	81	118
1936	98	83½	117½

Column (4) gives a measure of the average wage of a manual worker in the United Kingdom, expressed in terms of goods and services. The "stickiness" of money wages, combined with the fall in prices of foodstuffs associated with the world depression in agriculture, has resulted in a substantial improvement in the position of the typical wage earner, provided he is in continuous employment. In interpreting these figures we should make a mental addition on account of the all-round improvement in the qualities and varieties of the goods purchasable and the increasing value of social services to which he is entitled *gratis*. On the other hand, some deduction would appear to be reasonable as a measure of compensation for the increasing noise, congestion, and other drawbacks incidental to the march of progress. We conclude that in so far as the facts are susceptible of statistical measurement, the working classes of 1937 are substantially better off than those of 1924.

[1] To convert the cost of living index from base July, 1914, to base December, 1924, we multiply by the factor 100 ÷ 181 = 0·5525, where the figure 181 represents the index as published for December, 1924.

CHAPTER XXIII

EMPLOYMENT

ADEQUATE statistics of unemployment have been available since November 1920 when the unemployment insurance scheme was extended so as to cover 11 000 000 persons representing more than one half the gainfully occupied population of the country Prior to this date reliance must be placed on the returns made by certain Trade Unions [1] which extend back some fifty years Since the War the unemployment problem has changed its character and these returns (which were discontinued after 1926) possess only an historical interest

It is not proposed to discuss the causes of unemployment in this country nor will any elaborate statistical analysis be attempted The conditions of the problem are so uncertain and so changeable that analyses become out of-date almost as soon as they are made Attention will be confined to the official sources of information available in order that the reader may be in a position to follow current literature and draw his own conclusions

Sources of Information

The Ministry of Labour publish a detailed monthly statement of the employment position This statement includes a verbal summary an unemployment chart covering the current and previous years and a number of detailed tables Specimens are given on pages 251 and 252

Employers' Returns

Returns of the employment position are obtained from principal industries upon a voluntary basis and details are published monthly Specimen returns for the Cotton Industry are given in Table 85

The numbers of workpeople employed represent the numbers covered by the returns received and not the total numbers employed in the various industries In the comparisons of numbers employed

[1] See *Twenty first Abstract of Labour Statistics* (Cmd 4625) pp 68 and 69 also pages 48–52 of the present work

TABLE 85

UNITED KINGDOM—COTTON INDUSTRY—EMPLOYMENT IN MAY, 1937

SUMMARY OF EMPLOYERS' RETURNS [1]

—	Number of Workpeople			Total Wages Paid to All Workpeople		
	Week ended 29th May, 1937	Inc. (+) or Dec. (−) on a Month Before	Inc. (+) or Dec. (−) on a Year Before	Week ended 29th May, 1937	Inc. (+) or Dec. (−) on a Month Before	Inc. (+) or Dec. (−) on a Year Before
DEPARTMENTS		Per Cent	Per Cent	£	Per Cent	Per Cent
Preparing	11,489	− 0·2	+ 1·4	19,623	− 1·4	+ 10·1
Spinning	23,054	− 0·1	+ 2·5	38,900	− 1·9	+ 12·9
Weaving	19,054	− 0·2	+ 2·4	34,914	+ 0·2	+ 13·1
Other	6,501	− 0·4	+ 3·7	14,707	− 0·9	+ 8·9
Total	60,098	− 0·2	+ 2·4	108,144	− 1·0	+ 11·9
DISTRICTS						
Ashton	4,473	− 0·8	− 5·1	7,531	− 5·7	− 1·1
Stockport, Glossop, and Hyde	5,216	− 0·7	+ 6·4	9,017	− 2·5	+ 12·8
Oldham	9,398	− 1·0	+ 0·5	18,057	− 2·7	+ 8·3
Bolton and Leigh	11,279	+ 0·3	+ 7·9	19,560	+ 0·7	+ 18·9
Bury, Rochdale, Heywood, and Todmorden	5,498	− 1·0	− 3·4	10,286	− 0·0	+ 6·7
Manchester	3,458	− 0·9	− 3·7	5,898	− 5·5	+ 3·0
Preston and Chorley	4,053	+ 0·4	+ 0·5	7,215	+ 2·6	+ 8·7
Blackburn, Accrington, and Darwen	3,074	+ 1·2	+ 9·4	5,826	+ 2·1	+ 23·5
Burnley and Padiham	3,143	+ 1·0	+ 6·0	6,767	+ 4·9	+ 20·4
Colne and Nelson	2,455	− 0·3	+ 0·3	5,668	− 0·4	+ 18·5
Other Lancashire Towns	3,093	+ 0·7	+ 3·1	4,055	− 6·7	+ 7·5
Yorkshire Towns	2,005	− 0·5	+ 0·7	3,483	− 1·3	+ 17·5
Other Districts	2,953	+ 0·7	+ 7·1	4,781	+ 0·8	+ 18·3
Total	60,098	− 0·2	+ 2·4	108,144	− 1·0	+ 11·9

[1] *Ministry of Labour Gazette*, June, 1937, p. 224.

UNITED KINGDOM—UNEMPLOYMENT CHART

PERCENTAGES UNEMPLOYED[1] AMONG PERSONS AGED 16–64 INSURED AGAINST UNEMPLOYMENT IN GREAT BRITAIN AND NORTHERN IRELAND

1937	————	1934	- - - - - - -
1936	————	1933	• • • • • •
1935	---·---·---	Mean for 1924–29	o–o–o–o–o–o

x The crosses indicate the minimum and maximum monthly percentages unemployed during the years 1921–1936 [2]

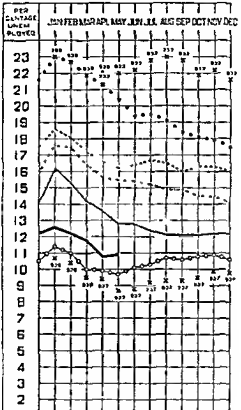


FIG 43

and wages paid at different dates the figures relate to the same firms at each date, and cover all the wage-earners, irrespective of age, sex, or occupation, employed by these firms. The precise information given in the employers' returns varies from industry to industry.

Unemployment in Insured Industries.

The Unemployment Insurance Acts provide, subject to certain exceptions, for the compulsory insurance against unemployment of substantially all employed persons. The principal classes of persons excepted are persons aged 65 and over, persons employed otherwise than by way of manual labour at a rate of remuneration exceeding in value £250[1] per annum, private domestic servants, and outworkers. Persons employed by local public authorities, railways and certain other public utility undertakings, members of the police forces, and persons with rights under a statutory superannuation scheme may, in certain circumstances, also be excepted. Statistics of (1) persons insured under the Agricultural Scheme and (2) juveniles under 16 years of age in insured industries are recorded separately.

An unemployment book, on which is recorded the industry in which he is employed, is issued to every insured person, and this book must be lodged at an Employment Exchange whenever the insured person makes a claim for unemployment benefit or for an unemployment allowance, or registers as unemployed without claiming benefit or allowances. The book must be removed and deposited with the employer for stamping as soon as employment in an insured trade is resumed.

The files of "lodged" books at the Employment Exchanges thus furnish for each industrial group a record of the unemployment of insured persons. In arriving at this figure the books of those persons who are known to be working in an uninsured trade, or to be sick or deceased, or to have gone abroad, are excluded. Where information on these points is lacking, the books remain in the "lodged" files, and are included in the statistics of unemployment, for a period of two months from the date on which the insured person was last in touch with the Exchange.

The numbers unemployed given in the following tables relate only to persons aged 16–64 insured against unemployment. They include insured persons maintaining registration at Employment

[1] Since altered to £420.

Exchanges together with those whose unemployment books remain lodged in the two months file referred to above

Insured persons who are disqualified for the receipt of unemployment benefit under the trade dispute disqualification are not included in the numbers unemployed unless they are definitely maintaining registration for other employment

The estimated numbers of insured persons in each industry are computed once a year in November on the basis of information obtained at the annual exchange of unemployment books The figures relate to the beginning of July and similar statistics are not available for other dates in the year In considering the figures on page 26. it should be borne in mind that in the case of individual industries the percentage rates of unemployment at April and May 1937 have been calculated on the basis of the estimated numbers of insured persons at the beginning of July 1936 while the figures for May 1936 and 1935 are calculated on the basis of the estimated numbers insured at July 1935 and 1934 respectively In an industry in which a relatively large change occurs during one or more years in the estimated number of insured work people this change may have an important effect on the relative percentage rates of unemployment

The following tables have been chosen from a large mass of material in order to illustrate the chief types of information available For details the student is referred to the original returns

Detailed Returns of Unemployment—Insured Industries

Table 86 shows the actual numbers recorded as unemployed in each industry whilst Table 87 shows the numbers insured and *percentages* unemployed at 24th May 1937

Numbers on Registers of Employment Exchanges

In addition to the figures referred to above the Ministry of Labour publishes monthly figures showing the total number of applicants for employment registered by the employment exchanges

Table 88 gives an analysis of these figures for 24th May 1937

The Ministry of Labour publishes a monthly summary showing the composition of the unemployment statistics which provides the necessary links in the chain of information (See Table 89)

TABLE 86

UNITED KINGDOM—NUMBER OF INSURED PERSONS RECORDED AS UNEMPLOYED AT 24TH MAY, 1937[1]

Industry	Great Britain and Northern Ireland									Great Britain only		
	Wholly Unemployed (including Casuals)			Temporary Stoppages			Total			Wholly Unemployed, Temporary Stoppages and Casuals		
	Males	Females	Total	Males	Females	Total	Males	Females	Total	Males	Females	Total
FISHING	8,455	118	8,573	127	35	162	8,582	153	8,735	8,527	153	8,680
MINING—												
Coal Mining	108,982	488	109,470	46,592	51	46,643	155,574	539	156,113	155,526	539	156,065
Iron Ore and Ironstone Mining, etc.	908	—	908	76	—	76	984	—	984	983	—	983
Lead, Tin, and Copper Mining	646	—	646	14	—	14	660	—	660	660	—	660
Stone Quarrying and Mining	4,616	12	4,628	682	—	682	5,298	12	5,310	4,339	11	4,350
Slate Quarrying and Mining	599	1	600	88	—	88	687	1	688	686	1	687
Other Mining and Quarrying	777	152	929	613	8	621	1,390	160	1,550	1,357	160	1,517
Clay, Sand, Gravel, and Chalk Pits	1,200	7	1,207	93	1	94	1,293	8	1,301	1,148	8	1,156
Total, Mining	*117,728*	*660*	*118,388*	*48,158*	*60*	*48,218*	*165,886*	*720*	*166,606*	*164,699*	*719*	*165,418*
NON-METALLIFEROUS MINING PRODUCTS												
Coke Ovens and By-product Works	1,596	6	1,602	131	—	131	1,727	6	1,733	1,726	6	1,732
Artificial Stone and Concrete	2,357	79	2,436	532	9	541	2,889	88	2,977	2,812	88	2,900
Cement, Limekilns, and Whiting	758	14	772	104	2	106	862	16	878	796	16	812
Total, N.-M. Mining Products	*4,711*	*99*	*4,810*	*767*	*11*	*778*	*5,478*	*110*	*5,588*	*5,334*	*110*	*5,444*
BRICK, TILE, PIPE, ETC., MAKING	6,672	560	7,232	691	93	784	7,363	653	8,016	6,947	653	7,600
POTTERY, EARTHENWARE, ETC.	3,033	2,053	5,086	1,839	3,133	4,972	4,872	5,186	10,058	4,842	5,183	10,025
[2]Total	1,012,093	203,496	1,215,589[2]	149,283	61,118	210,401	1,191,376	264,614	1,455,990	1,144,564	245,776	1,390,340

[1] *Ministry of Labour Gazette*, June, 1937, p. 228.
[2] Excluding persons insured under the agricultural scheme and juveniles under 16 years of age.

TABLE 88

NUMBERS ON THE REGISTERS OF EMPLOYMENT EXCHANGES[1]

ANALYSIS FOR 24TH MAY, 1937, AND 19TH APRIL, 1937

—	24th May, 1937				19th April, 1937
	Persons Normally in Regular Employment		Persons Normally in Casual Employment	Total	Total
	Wholly Unemployed	Temporarily Stopped			
Great Britain					
Men . . .	939,438	144,230	68,199	1,151,867	1,141,011
Boys. . .	32,056	4,197	167	36,420	41,315
Women . .	167,299	55,491	1,925	224,715	227,026
Girls . . .	35,009	3,310	9	38,328	45,091
Total . .	1,173,802	207,228	70,300	1,451,330	1,454,443
Great Britain and Northern Ireland					
Men . . .	982,805	145,479	70,838	1,199,122	1,191,632
Boys. . .	34,206	4,204	167	38,577	43,562
Women . .	182,419	57,942	1,955	242,316	244,507
Girls . . .	35,820	3,416	9	39,245	46,007
Total . .	1,235,250	211,041	72,969	1,519,260	1,525,708

[1] *Ministry of Labour Gazette*, June, 1937, p. 221.

Insured Population.

Estimates of the numbers of persons aged 16–64 insured under the Unemployment Insurance Acts, classified by industries, are available annually for July of each year. Separate figures are given for males and females and for Great Britain and the United Kingdom.[1] The figures for the United Kingdom are re-published in the form of index numbers (July, 1923 = 100), the tables being conveniently arranged so as to distinguish expanding from contracting industries. As all persons aged 65 and over ceased to be insured under the Acts as from 2nd January, 1928, there is a break in the continuity of the figures, which has been smoothed out in the manner illustrated below—

		Thous.
Estimated number of insured persons aged 16 and over—Great Britain	1923	11,150
	1927	11,750
Estimated number of insured persons aged 16 to 64 —Great Britain	1927	11,408
	1928	11,500
1927 as percentage of 1923—11,750/11,150 × 100 =		105·4
1928 as percentage of 1927—11,500/11,408 × 100 =		100·8
1928 as percentage of 1923 (allowing for change in population at risk)—1·054 × 100·8 × 100 =		106·2

The numbers of the insured population at dates intermediate between successive Julys are not known precisely. Methods of estimating are discussed in the next section.

Insured Persons in Employment.

The *Ministry of Labour* publishes a monthly estimate of the total number of insured persons aged 16–64 exclusive of those in the agricultural scheme, in employment in Great Britain, compiled by adding the total number of entrants into insurance each month (a known quantity) and deducting the number of exitants (which is not known precisely and must be estimated). The difference represents the total number not recorded as unemployed. See Table 90, column (4).

The figures in column (6) have been obtained by deducting from the total estimated numbers insured the numbers recorded as unemployed and the numbers directly involved in trade disputes, together with an allowance of $3\frac{1}{2}$ per cent of the numbers insured in respect of absences from work through sickness and other forms of unrecorded non-employment apart from "recognized" holidays.

[1] *Twenty-first Abstract of Labour Statistics* (Cmd. 4625) and *Ministry of Labour Gazette*

The estimated numbers insured, and numbers in employment, from August, 1936, onwards are provisional, and subject to revision when information as to the numbers of insured persons becomes available from the annual exchange of books in 1937.

The Ministry also publishes annual indices of employment[1] by industry on the same lines as the indices of insured population described in the previous section. In this case, however, no account is taken of sickness, trade disputes, etc.

No monthly figures are available for employment in individual industries. For industries not subject to rapid expansion or contraction it is possible to make estimates on the lines indicated below.

TABLE 91

ESTIMATED EMPLOYMENT IN THE BUILDING INDUSTRY—UNITED KINGDOM[1]

July, 1935–*May*, 1937

(Thousands of Persons)

Year and Month (1)	No. Insured (2)	No. Unemployed (3)	Balance = No. Employed (4)
1935—			
7	977	140	837
8	981	137	844
9	984	144	840
10	988	148	840
11	991	165	826
12	995	178	817
1936—			
1	998	274	824
2	1,002	198	804
3	1,005	145	860
4	1,009	125	884
5	1,013	108	905
6	1,016	109	907
7	1,020	117	903
8	1,022	115	907
9	1,024	124	900
10	1,026	135	891
11	1,028	154	874
12	1,030	180	850
1937—			
1	1,032	183	849
2	1,034	167	867
3	1,036	166	870
4	1,038	122	916
5	1,040	118	922

[1] *A Twenty-first Abstract of Labour Statistics* (Cmd. 4625), pp. 43–4.

The only figures in column (2) known exactly are those for July 1935 and 1936. The rest of the column is found by linear interpolation up to July 1936 and thereafter by extrapolation.[1] Column (3) is known exactly and column (4) is given by difference. Owing to differences in definition the figures obtained by this method may differ considerably from those given in the Census of Production Statistics.

Comparability

The above figures illustrate some of the difficulties experienced in the measurement of social phenomena. The *desideratum* is a satisfactory measure of unemployment, i.e. inability of people who usually follow gainful occupations and are not temporarily or permanently disabled from following them, to find work suitable for their capacities.

Ideally, the statistics should cover the whole of the gainfully occupied population employed in subordinate capacities. The insured population, which forms the basis of the Ministry's calculations, covers only a substantial fraction of this field.

Persons who register at the employment exchanges do so for one or more of the following reasons—

1. In order to qualify for benefits or transitional payments.
2. For assistance in obtaining employment.
3. In order to have their health insurance cards franked.
4. As a condition for receipt of public assistance.

Naturally, changes in legislation and administrative regulations vary inducements to lodge books or apply at the exchanges from time to time. Complications are also introduced by the fact that some applicants for employment are not insured persons. These factors affect the comparability of the statistics. Taking the broad view, however, the picture given by the Ministry's figures is sufficiently accurate for most practical purposes.

[1] For the period July 1935 to July 1936 we add 1/12 (1020 − 977) = 3·583 for each successive month. For the period subsequent to July 1936 we add 0·2 per cent (simple interest) for each successive month, upon the provisional assumption that the insured population in the building industry is increasing at the same rate as that of industry in general. When the true figure for 1937 becomes available (in November 1937) this section of the table will require re-calculation.

CHAPTER XXIV

PROFITS

Inland Revenue Returns.

STATISTICS of assessments under *Schedule D* (profits from businesses, professions, and certain interest) are published annually by the Board of Inland Revenue. The following table gives a summary for the fiscal year 1933–34.

TABLE 92

INCOME TAX, SCHEDULE D: UNITED KINGDOM—PROFITS FROM BUSINESSES, PROFESSIONS, AND CERTAIN INTEREST

(Assessments made in 1933–34)[1]

		£(million)
(a) Gross Income—		
Manufacturing, Productive, and Mining Industries		332
Distribution, Transport, and Communication—		
Railways		24
Other Assessments		367
Finance Professions and Other Profits		177
Interest		97
Dominion and Foreign Securities and Possessions		73
		£1,070
(b) Exemptions—	£(million)	
Incomes below the effective exemption limit	8	
Charities, etc.	21	
Dominion or Foreign Dividends belonging to Non-residents	2	
(c) Reductions—		
Allowances for Wear and Tear	96	
Other Reductions and Discharges	187	
		314
(d) Actual Income, viz. (a) less (b) and (c)		756

These figures give satisfactory indications of the annual movement of business profits, but they need interpretation by an expert. Statutory income is not identical with income of the current year, adjustments must be made for evasions and losses not included in

[1] *Seventy-eighth Report of the Commissioners of Inland Revenue* (Cmd. 5015), 1936, p. 69. Statistics for past years are given in the same publication.

the figures, whilst the actual allowances for depreciation and wear and tear are probably insufficient

It is also unfortunate that the Board of Inland Revenue are unable to analyse the reductions under the same headings as the gross income

The "Economist's" Statistics of Industrial Profits.

The *Economist* publishes monthly, quarterly and annual statistics of industrial profits, compiled on a sample basis

Monthly Statistics All reports of public companies registered in the United Kingdom published during the current month are tabulated and their aggregate net profits (after deduction of debenture interest) are ascertained and compared with the profits of the *same* companies for the preceding year (any companies not represented in both years being struck out for the purposes of comparison) For June, 1937, the number of reports brought into the calculation was 265 the total net profits being £43 924,563, compared with £36 244 177 for the preceding year The rise in profits was therefore 21 19 per cent Table shows the results for the period March, 1936, to June 1937

Quarterly Statistics. These are compiled upon the same principles but on more elaborate lines. Table 94 gives a specimen.

The chain index is constructed as follows—

$$1931/1930 \quad . \quad . \quad . \quad \frac{677}{758} \times 100 = 89{\cdot}3$$

$$1932/1930 \quad . \quad . \quad . \quad \frac{543}{634} \times 89{\cdot}3 = 76{\cdot}5$$

The rest of the table is self-explanatory.

Annual Statistics. These are compiled on the same lines as the quarterly statistics. The chain index now expresses the relation between successive years instead of individual quarters of successive years.

Comments.

These figures represent the only continuous record of movements of profits by industries that is available. Let us examine their construction more closely.

The monthly statistics proceed by sample. If the samples were truly representative in the sense that every industrial concern had

TABLE 94

STATISTICS OF INDUSTRIAL PROFITS—MARCH QUARTER, 1937

I. Industrial Profits Index

("First Quarter" Companies)

Published during First Quarter of	No of Companies	Net Profits (after Debenture Interest)		*Chain Index (1930 = 100)*
		Year Stated	Same Companies preceding Year	
		£	£	
1930	—	—	—	*100 0*
1931	596	67,736,226	75,816,474	*89·3*
1932	548	54,314,815	63,403,727	*76·5*
1933	562	52,046,133	57,105,291	*69·7*
1934	547	55,751,307	52,922,902	*73·4*
1935	592	67,137,496	58,577,585	*84·1*
1936	569	75,731,976	66,839,644	*95·3*
1937	628	94,733,138	84,661,458	*106·7*

II Net Profits in Individual Industries

(After payment of Debenture Interest etc)

Group	No of Reports	Reports Published in Quarter ended 31st March: 193	Compared with 1936	+ or −	No of Reports	Reports Published in Nine Months ended 31st March: 1937	Compared with 1936	+ or −
				%		£	£	%
Breweries	15	1 9 537	+ 90 421	4 8	87	15 868 905	+ 1 252 564	8 6
Canals and docks	6	437 784	+ 14 889	3 5	6	457 784	+ 14 889	5 5
Electric lighting and power	24	9 879 343	910 929	10 1	28	10 515 707	+ 1 025 725	10 8
Financial land and investment	47	6 57 303	+ 1 6 375	36 2	136	11 192 950	+ 2 577 587	26 9
Gas	31	3 267 217	36 559	1 1	34	5 581 228	45 530	1 5
Hotels and restaurants	4	78 81	9 152	15 1	17	604 575	+ 55 207	9 6
Iron coal and steel	34	6 707 721	1 679 126	55 5	101	15 193 697	+ 4 035 764	36 1
Motors cycles and aviation	7	1 470 500	+ 155 918	11 8	35	5 95 305	+ 1 232 488	27 0
Oil	1	1 9 55	+ 37 101	46 8	9	1 405 386	+ 429 830	44 1
Rubber	62	07 130	+ 571 008	110 5	254	2 215 782	+ 519 783	30 7
Shipping	6	948 915	+ 175 885	22 7	14	5 065 546	+ 63 740	2 1
Shops and stores	42	10 024 758	+ 727 829	7 8	64	15 060 725	+ 1 031 505	8 6
Tea	4	38 197	+ 6 807	21 7	49	745 944	+ 406 846	119 9
Telegraphs and telephones	3	365 253	+ 45 909	14 4	3	563 253	+ 45 909	14 4
Textiles	30	4 198 305	+ 350 492	9 1	60	5 643 375	+ 501 609	9 7
Tramway and omnibus	9	1 314 696	+ 262 147	24 9	16	1 248 798	+ 203 313	19 4
Trusts	69	5 552 069	+ 514 285	9 7	142	6 144 628	+ 437 340	7 7
Waterworks	20	766 102	16 151	2 0	20	809 139	15 394	1 9
Building materials	28	4 399 904	+ 398 448	9 9	53	6 462 527	+ 657 412	11 3
Food confectionery and drink	23	1 828 625	+ 107 857	6 3	55	6 319 025	+ 127 569	2 0
Electrical equipment		2 664 677	+ 390 867	28 5	14	5 305 130	+ 714 474	27 6
Newspapers, printing, etc	18	2 218 030	+ 366 078	29 7	42	5 691 931	+ 584 055	18 8
Tobacco	5	14 370 133	+ 742 770	5 5	15	21 617 771	+ 1 073 572	5 2
Warehouse and trading	15	468 327	+ 43 415	10 1	25	1 054 359	+ 125 720	13 3
Miscellaneous	115	16 599 679	+ 1 040 514	6 7	246	25 01 011	+ 2 404 743	10 5
Total	628	94 733 138	+10 071 680	11 9	1503	165 842 475	+ 19 256 336	13 1

III. Percentage Comparison with Preceding Years

	1930–31	1931–32	1932–33	1933–34	1934–35	1935–36	1936–37
Third quarter .	− 6·4	− 35·5	− 28·6	− 5·5	+ 27·7	+ 12·2	+ 14·3
Fourth quarter .	− 18·1	− 53·9	− 2·9	+ 30·3	+ 32·9	+ 16·9	+ 15·2
First quarter .	− 10·6	− 14·3	− 8·9	+ 5·3	+ 14·6	+ 13·3	+ 11·9
Second quarter .	− 19·4	− 21·8	+ 3·3	+ 18·2	+ 17·8	+ 10·6	—
Year ending 30th June	− 14·7	− 26·5	− 6·7	+ 12·1	+ 19·5	+ 12·6	—

IV. Ordinary Share Ratios

Companies Reporting in First Quarter of	Total Ordinary Capital	Earned for Ordinary	Paid on Ordinary	"Ploughed Back" (Ratio to Ordinary Capital)
	(£'000)	%	%	%
1930 . .	383,642	13·0	10·0	3·0
1931 . .	494,000	10·7	8·5	2·2
1932 . .	430,189	9·9	8·7	1·2
1933 .	443,645	8·9	7·5	1·4
1934 . .	441,919	9·8	7·5	2·3
1935 . .	483,029	11·0	8·3	2·7
1936 . .	484,890	12·6	9·2	3·5
1937 . .	527,503	14·4	10·2	4·2

V. Distribution of Profits

	Net Profits after Payment of Debenture Interest	Ordinary Dividend		Preference Dividend		To Reserve, Etc.	
1936—	£	£	£	£	%	£	%
First quarter .	75,731,976	44,481,431	58·7	14,457,695	19·1	16,792,850	22·2
Second quarter .	92,280,660	47,225,745	51·2	20,527,591	22·2	24,527,324	26·6
Third quarter .	28,974,570	14,647,169	50·5	5,728,751	19·8	8,598,650	29·7
Fourth quarter .	42,134,767	22,557,928	53·5	7,390,205	17·5	12,186,634	29·0
Total 1936	£239,121,973	£128,912,273	53·9	£48,104,242	20·1	£62,105,458	26·0
Total 1935	£203,246,989	£111,720,344	54·9	£46,278,011	22·7	£45,248,634	22·4
1937— First quarter .	94,733,138	53,609,355	56·6	18,775,905	19·8	22,347,878	23·6

II Average Rates Paid

	On Debenture Capital				On Preference Capital				On Ordinary Capital			
	1934	1935	1936	1937	1934	1935	1936	1937	1934	1935	1936	1937
	%	%	%	%	%	%	%	%	%	%	%	%
First quarter	4·71	4·62	4·51	4·30	4·9	5·1	5·3	5·0	7·5	8·3	8·2	10·2
Second quarter	4·73	4·33	4·35	—	5·2	5·7	5·6	—	5·7	6·7	7·3	—
Third quarter	4·83	4·67	4·51	—	4·1	4·9	4·9	—	6·0	6·7	7·6	—
Fourth quarter	4·95	4·84	4·75	—	3·2	4·5	4·9	—	6·8	7·9	9·2	—
Year	4·82	4·72	4·62	—	4·8	5·2	5·2	—	6·5	7·4	8·2	—

of the total area of industry than did the previous year's. In view of our present rationalization programme, it must be assumed that the area covered is on the increase, and to the extent that this is so, will any percentage increase in profits be inflated.

There is a lag in the figures due to the amount of time necessary to make up and publish the accounts at the end of each year.

Weighing up the evidence, it appears that the *Economist's* figures, whilst giving a correct indication of the direction of profit movements, suffer from an upward bias due to the factors in question.

Profits and Capital.[1]

Attempts to measure returns upon invested capital are not statistically satisfactory. It is extremely difficult to say what is the amount of capital invested in a business, and answers to the question will be different according to the standpoint adopted.

Capital may be distinguished as **Owned Capital** and **Borrowed Capital**. Under modern joint-stock enterprise owned capital is usually represented by equity shares and borrowed capital by debentures and permanent loans. Short-term loans, bank overdrafts, and similar accommodation are not usually regarded as forms of capital.

The most natural method of approaching the problem of return upon capital is to go back to the formation of the business and to tabulate the amounts put in by the various providers of capital, including undistributed profits, deducting any amounts paid off. The result represents the amount of investment in the business, it being assumed, of course, that adequate allowance for depreciation has been made before profits are struck, and it would be represented in the balance sheet by the sum total of loan capital, share capital, capital reserves, and free reserves.

In course of time this figure becomes divorced from the value of the business as a going concern. From an asset standpoint the values of capital goods change, and provision for obsolescence must be made. From the earnings standpoint, there will be changes in the amount of profits, whilst variations in long-term interest rates and in the prospects of business will affect the basis of capitalization (i.e. the number of years' purchase). The new investor will not be interested in past history, he will study the present and future as reflected in the market values of the company's shares. And since

the ownership of a modern company is always changing hands there is no definite figure upon which to base calculations of current return Every investor will calculate his return in his own way and the nearest approach to stability will be given by the yield on the current price of the shares assuming the position of the market is technically sound

Sir Josiah Stamp's Index of Profits

Sir Josiah Stamp has constructed a new and improved series (1) by regrouping the *Economist's* figures so as to bring them as nearly as possible into line with the calendar year and (2) by averaging the series with another based on the Inland Revenue returns The results are published annually in the form of (1) a General Index showing the fluctuations in the return to Industrial Capital including debentures and other relatively immobile yields and (2) a Special Index showing the return on ordinary shares and equity risks[1] The index is continued annually in the correspondence columns of *The Times* The figures for the period 1920–36 are as follows—

TABLE 96

PROFIT AS PERCENTAGE OF TURNOVER—AGGREGATE OF SEVEN INDUSTRIAL GROUPS

Value of the Variable (1)	1912–13 Percentage of the Total Turnover (2)	1922–23 Percentage of the Total Turnover (3)
– 20 and below	0·03	0·87
– 19 9 to – 10	0·03	2·07
– 9·9 „ – 5	0·04	2·39
– 4·9 „ ·0	1·92	8·96
0 „ ·9	9·44	8·33
1 „ 1·9	7·69	9·86
2 „ 2·9	11·02	10·25
3 „ 3·9	14·57	6·45
4 „ 4·9	8·69	7·41
5 „ 5·9	11·45	7·48
6 „ 6·9	6·22	5·52
7 „ 7·9	5·86	3·85
8 „ 8 9	3·96	3·40
9 „ 9 9	4·01	2·85
10 „ 10·9	2·23	2·98
11 „ 11·9	2·59	1·75
12 „ 12·9	2·95	1·29
13 „ 13·9	1·48	1·41
14 „ 14·9	0·54	1·08
15 „ 19 9	3·24	5·47
20 „ 24·9	1·10	3·31
25 „ 29·9	0·18	1·51
30 „ 39 9	0·59	0·93
40 „ 49 9	0 05	0·33
50 and over	0·12	0·25
	100·00	100·00

	1912–13	1922–23
Median	4·61	4·11
Lower quartile	2·53	1·24
Upper quartile	7·67	8·46
Skewness	+ 0·19	+ 0·20
Mean	5·80	5·43
Mean deviation from median	3·59	6·01

Profits Upon Turnover.

As an alternative to calculating profits upon capital they may be calculated upon turnover The figures on p 271 are extracted from details furnished to the *Committee on National Debt and Taxation* [1]

Retail Profits

Little is known on the subject of retail profits and such information as is available suggests that there are wide discrepancies between different types of business

[1] For further details see the Memorandum by Dr W H Coates presented to the *Committee on National Debt and Taxation* (1927) Appendix XI p 65 *seq*

CHAPTER XXV

TRADE

1. Overseas Trade

Statistics of Overseas Trade are collected and published in considerable detail. These include—

1. Monthly "Accounts," issued on the twelfth working day of the month following that to which they relate. The summary tables are reproduced in the daily Press.

2. Quarterly summaries published in the *Board of Trade Journal.*

3. Annual "Statements," published in four volumes, at considerable intervals after the close of the year to which they refer.

Scope and Definitions.

The Monthly Accounts show quantities and values of principal articles imported and exported for the current month, the expired portion of the current year and corresponding periods of the previous two years.

Particulars are given for every item specified in the official import and export list.

The tables are arranged in three sets, showing—

A. Total Imports.

B. Exports of Produce and Manufactures of the United Kingdom.

C. Exports of Imported Merchandise.

Deduction of C from A gives figures of Net Imports (Imports retained in the United Kingdom). Subject to adjustments for changes in stocks, statistics of Net Imports give a measure of Home Consumption of imports of the article in question.

In the absence of special qualification, statistics of "Imports" and "Exports" are usually understood to refer to A and B respectively.

Values of Imports are reckoned c.i.f. (cost, insurance freight) and values of exports f.o.b. (free on board). This practice is followed

by most other countries and in consequence the value of total imports for the whole world tends to exceed that of total exports by about 6 per cent the difference representing values added in transit

Imports are classified as received from the place or country of consignment (This is not necessarily the place or country of shipment origin or manufacture) Exports are classified as dispatched to the country of final destination (This is not necessarily the place or country of unshipment)

The monthly returns are elaborated in the quarterly summaries and further elaborated in the Annual Statements Here we find in addition classifications by countries and ports Convenient summaries of the Annual Statements are given in the *Statistical Abstract for the United Kingdom*

It should be pointed out that the results of adding up the twelve monthly figures usually differ somewhat from the annual figures as finally published The reason for this discrepancy is that the final figures embody a number of corrections which are not carried back into the monthly figures on account of the excessive amount of work that would be involved

Certain classes of goods taken into or out of the country are excepted from the returns For particulars see the original publications

Visible Trade Balance

The scope of these returns is best understood by reference to Table 97 showing the visible trade balance of the United Kingdom for the years 1934 36

The figures for imports and exports represent debits and credits on current account and may not correspond with payments and receipts for the Calendar Year Movements of gold bullion and specie are regarded as connected with capital transactions

Classification of Commodities

The five main heads of classification of commodities are shown below The figures in brackets represent the numbers of subclasses

I Food drink and tobacco (9)

II Raw materials and articles mainly unmanufactured (14)

TABLE 97

UNITED KINGDOM—VISIBLE TRADE BALANCE, 1934–36[1]

Movement	1934	1935	1936
Imports—	£ (million)	£ (million)	£ (million)
Merchandise	731·4	756·0	848·9
Silver bullion and specie . . .	22·2	40·5	17·1
Total	753·6	796·5	866·0
Exports—			
Merchandise	447·2	481·1	501·1
Silver bullion and specie . . .	12·5	54·9	18·1
Total	459·7	536·0	519·2
Excess of Imports over Exports . .	293·9	260·5	346·8

III. Articles wholly or mainly manufactured (20).
IV. Animals not for food.
V. Parcel Post (non-dutiable articles).

The summary figures for the period 1924 to 1936 are given in Table 98.

Volume and Value of Overseas Trade.

The figures in Table 98 are expressed in sterling and are likely to convey a misleading impression of the fluctuation of the actual volume of this country's external trade. Accordingly, the Board of Trade publishes a series of figures in which trade for the current quarter is revalued in terms of average prices for 1935. The procedure may be illustrated as follows—

	£ (million)
(*A*) Value of goods for current year as declared . . .	300
(*B*) Value of goods included in above for which definite unit prices are available	200
(*C*) Value of goods included under (*B*) at prices of base year, 1935	250

Estimated value of current year's goods at base year prices

$$= \frac{250}{200} \times 300 = £375{,}000{,}000$$

[1] See *Board of Trade Journal*, 25th February, 1937.

TABLE 98 (*continued*)

Year	Class I: Food, Drink, and Tobacco			Class II Raw Materials, etc			Class III Articles wholly or mainly Manufactured			All Classes, including IV and V		
	British Countries	Foreign Countries	Total	British Countries	Foreign Countries	Total	British Countries	Foreign Countries	Total	British Countries	Foreign Countries	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Exports of Imported Merchandise												
1924	12·8	17·1	29·8	1·9	74·3	76·2	11·8	22·0	33·7	26·5	113·5	140·0
1925	13·3	18·8	32·1	2·1	88·3	90·3	10·9	20·5	31·5	26·3	127·7	154·0
1926	13·1	13·2	26·4	1·9	71·9	73·8	8·6	16·6	25·2	23·7	101·8	125·5
1927	11·9	14·6	26·5	1·5	69·7	71·2	8·6	16·4	25·0	22·0	100·9	123·0
1928	12·6	14·9	27·5	1·6	64·8	66·4	8·4	17·6	26·0	22·8	97·5	120·3
1929	12·5	13·5	26·0	1·3	53·0	54·3	8·9	20·0	28·9	23·1	86·6	109·7
1930	11·2	12·6	23·8	1·3	37·1	38·4	7·7	16·4	24·1	20·5	66·4	86·8
1931	9·2	10·9	20·1	0·9	24·8	25·7	5·6	11·8	17·4	16·1	47·8	63·9
1932	6·9	8·2	15·2	0·8	22·9	23·7	3·6	8·3	11·8	11·5	39·5	51·0
1933	6·2	5·9	12·1	0·9	24·7	25·6	3·2	7·9	11·1	10·4	38·6	49·1
1934	6·7	6·0	12·6	1·2	26·5	27·7	3·1	7·6	10·6	11·1	40·2	51·2
1935	6·2	6·4	12·6	1·2	27·7	29·0	3·4	10·1	13·5	11·0	44·3	55·3
1936	[2]	[2]	11·7	[2]	[2]	33·0	[2]	[2]	15·4	[2]	[2]	60·4

[1] *Statistical Abstract for the United Kingdom* 1922–35 (Cmd. 5353), pp. 374–5.
Not yet available

the base period—new trades arise and the relative importance of old-established trades varies—and apart from the fact that new headings may be raised in the trade accounts which had no counterpart in the base year, the difference in weighting of the various items may affect materially the results of the revaluation. Accordingly, the Board have decided to change the base year to 1935, the year in respect of which the fifth Census of Production was taken.

Balance of Payments of the United Kingdom.

The published foreign trade figures of this country relate solely to imports and exports of tangible goods, bullion, and specie, and in order to show the true position it is necessary to take into account payments for services rendered as well as profit and interest upon capital loaned in the past and current capital movements.

These items form the subject of an annual estimate by the Board of Trade. The figures for the years 1934–36 are given in Table 100 below.

TABLE 100

BALANCES OF CREDITS AND DEBITS IN THE TRANSACTIONS (OTHER THAN THE LENDING AND REPAYMENT OF CAPITAL) BETWEEN THE UNITED KINGDOM AND ALL OTHER COUNTRIES, 1934–36

Particulars	1934	1935	1936
	In Million £'s		
Excess of imports of merchandise and silver bullion and specie	294	260	347
Estimated excess of Government payments made oversea	—	2	2
Total	294	262	349
Estimated excess of Government receipts from oversea[1]	7	—	—
Estimated net national shipping income[2]	70	75	95
Estimated net income from oversea investments	170	180	195
Estimated net receipts from commissions, etc.	30	30	30
Estimated net receipts from other sources	10	10	10
Total	287	295	330
Estimated total credit or debit balance on items specified above	−7	+33	−19

[1] Including some items on loan accounts.
[2] Including disbursements by foreign ships in British ports

The balance of Government receipts provides for receipts and payments on account of reparations and inter governmental loans (principal and interest) With this exception the table refers to revenue as opposed to capital transactions The balances of credit (or debit) therefore represent imports or exports of capital [1]

2 Internal Trade

By arrangement with the Incorporated Association of Retail Distributors and the Bank of England statistics relating to retail trade in Great Britain are published monthly in the *Board of Trade Journal* These sales figures are compiled from schedules issued by the Retail Distributors Association the Co operative Union the Drapers Chamber of Trade of Great Britain and Ireland the London Furniture Trades Federation and the Shoe Distributors Association to their members and to other traders who have agreed to collaborate among whom members of the Federated Multiple Shop Proprietors are represented they relate to the trade of a number of department stores concerns operating multiple retail shops independent retailers and a representative section of the retail co operative societies

Comparability of Published with Individual Statistics

In comparing sales figures for individual concerns with the published results it should be remembered that retailers use a variety of accounting periods and also that the number of selling days in any calendar month varies from year to year Returns have therefore to be adjusted to some extent in order to render the sales for 1936 and 1937 comparable when combined into district or other totals Almost all contributors now report the number of days on which selling took place in their shops during the period to which their figures relate On each return therefore the 1936 sales are corrected where necessary by the appropriate amount to make them comparable with those of 1937 so far as length of period is concerned Adjustments of this kind may be imperfect where a return relates to the sales of a number of shops in different towns but the error involved is not likely to affect the published figures Corrections for such differences between April 1936 and April

[1] See *Board of Trade Journal* 25th February 1937 which gives full particulars of the bases upon which these estimates are made

TABLE 101

RETAIL TRADE AS COMPARED WITH A YEAR AGO

APRIL AND FEBRUARY—APRIL; COMPARISON OF 1937 WITH 1936. (The figures shown are the percentage changes)

I SALES (AT SELLING VALUE) AND STOCKS (AT COST)

Class of Merchandise	Sales (on an approximate Daily Basis)									Stocks
	April								*Feb.–April*	At End of April
	Scotland	North East	North West	Midlands and South Wales	South of England	*London-*Central and West End	London-Suburban	Total, Great Britain	*Total, Great Britain*	Total, Great Britain
Piece-goods[1]	+ 1·1	+ 13·6	+ 10·8	+ 7·6	+ 6·0	+ 15·7	+ 7·3	+ 11·2	*+ 6·3*	+ 1·0
(i) Household goods	+ 3·8	+ 12·1	− 1·8	− 2·2	+ 2·8	+ 13·1	+ 2·3	+ 5·3	*+ 6·9*	+ 9·4
(ii) Dress materials	+ 1·4	+ 16·0	+ 17·0	+ 9·8	+ 9·0	+ 16·5	+ 9·0	+ 14·1	*+ 6·0*	− 3·7
Women's wear[1]	+ 3·0	+ 1·9	+ 1·6	+ 4·4	+ 2·6	+ 9·6	− 3·5	+ 4·3	*+ 8·8*	+ 6·5
(i) Fashion departments	+ 7·0	+ 9·1	+ 11·5	+ 14·5	+ 7·9	+ 17·1	+ 10·4	+ 12·5	*+ 13·1*	+ 7·1
(ii) Girls' and children's wear	+ 7·0	+ 9·5	− 2·0	+ 17·0	+ 7·9	− 7·8	+ 0·6	+ 1·2	*+ 8·5*	+ 3·4
(iii) Fancy drapery	− 1·8	− 6·9	− 7·8	− 7·7	− 6·2	+ 4·2	− 13·7	− 5·3	*+ 5·1*	+ 6·4
Men's and boys' wear	+ 5·2	+ 1·9	+ 3·0	+ 7·3	+ 6·5	− 1·3	− 2·0	+ 2·9	*+ 6·8*	+ 4·5
Boots and shoes	− 1·9	− 3·6	− 6·2	+ 2·5	+ 0·4	+ 0·6	− 9·2	− 2·8	*+ 10·7*	+ 2·3
Furnishing departments	− 0·7	− 1·7	+ 2·6	+ 1·4	− 1·0	+ 13·0	+ 2·2	+ 4·6	*+ 7·7*	+ 8·7
Hardware	+ 8·0	− 8·0	− 12·7	− 4·7	− 4·5	+ 15·4	+ 1·0	+ 0·6	*+ 1·6*	+ 9·8
Fancy departments	− 4·9	− 2·7	− 3·7	− 2·1	− 6·1	+ 0·4	− 5·0	− 3·6	*+ 2·7*	+ 1·3
Sports and travel	− 6·1	− 13·9	− 24·5	− 13·7	− 16·8	− 3·6	− 31·1	− 11·2	*− 0·4*	+ 10·2
Miscellaneous and unallocated	− 2·7	+ 0·5	+ 2·1	− ·08	− 3·8	+ 39·9	− 8·4	+ 2·0	*+ 11·4*	+ 16·8
Total of above	+ 0·8	+ 0·1	+ 0·2	+ 2·4	− 0·4	+ 9·4	− 3·1	+ 2·2	*+ 7·6*	+ 5·6
Grocery, provisions and bakery	− 0·3	+ 2·6	+ 0·8	+ 2·1	− 0·4	− 2·1	+ 0·7	+ 0·8	*+ 8·1*	+ 4·4
Other food and perishables	− 2·4	*Nil*	− 0·8	— 1·1	− 0·4	− 5·9	+ 0·2	− 0·8	*+ 7·9*	+ 16·9
Total—Food and Perishables[1]	*Nil*	+ 2·3	+ 0·3	+ 1·0	+ 0·5	− 3·6	+ 0·6	+ 0·8	*+ 8·1*	+ 6·5
TOTAL SALES—April	+ 0·3	+ 1·3	+ 0 2	+ 1·9	+ 0·2	− 8·0	+ 0·7	+ 1·5		
February-April	*+ 6·1*	*+ 8·0*	*+ 7·1*	*+ 9·1*	*+ 7·7*	*+ 8·2*	*+ 8·4*		*+ 7·9*	
TOTAL STOCKS—April	+ 2·0	+ 5·7	+ 4·3	+ 3·2	+ 1·7	+ 7·5	+ 6·1			+ 4·8

[1] Including some goods for which separate particulars under the sub-headings are not available.

II Persons Employed in April 1937 as Compared with April 1936

	Scotland	North East	North West	Midlands and South Wales	South of England	London Central and West End	London Suburban	Total, Great Britain
Total employees	+ 35	+ 41	03	+ 41	+ 26	+ 57	+ 09	+ 31
Selling employees	+ 09	+ 59	+ 08	+ 52	+ 53	+ 58	+ 42	+ 41
Juveniles (under 18 years of age)[1]	+ 197	+ 146	+ 70	+ 224	+ 126	+ 130	+ 112	+ 134

[1] Whether engaged in selling or not

III Sales in London and Rest of Great Britain as Compared with 1936

Class of Merchandise	Three Months February–April: London (Central West End and Suburban)	Three Months February–April: Rest of Great Britain
Piece-goods	+ 87	+ 40
Women's wear	+ 92	+ 85
Men's and boys' wear	+ 44	+ 78
Boots and shoes	+ 82	+ 114
Furnishing departments	+ 91	+ 65
Hardware	+ 71	− 19

Class of Merchandise	Three Months February–April: London (Central West End and Suburban)	Three Months February–April: Rest of Great Britain
Sports and travel	+ 48	− 71
Food and perishables	+ 83	+ 81
Total sales (including above and other kinds of merchandise)	+ 83	+ 77

IV—INDEX NUMBERS OF RETAIL SALES

(Average Daily Sales in 1933 = 100)

APRIL, 1937

—	Scotland	Provincial—England and Wales	London—Central and West End	London—Suburban	Total, Great Britain
Food and perishables	118	123	93	128	122
Other merchandise	109	117	117	106	115
Total sales	114	120	114	120	119

1937 were very numerous owing to the Easter holiday falling in April 1936 and in March 1937

It must also be remembered that the figures refer as far as possible only to the trade of branches or departments which have been established for a year at least and which were therefore actually in operation in the two months for which the return is made The object of this is to minimize the danger of showing as a general expansion of sales in any district movements which are not representative of average operating results and which possibly reflect transfers of trade from established to newly opened concerns Once the new branch or department can report its own sales not only for the month in the current year but also for the same month a year earlier the percentage movements recorded should not be seriously affected but as the first comparisons will be with a period of expanding sales the published statistics will have a favourable bias Further the difficulty of ensuring adequate representation of the smaller individual trader in these statistics must always be borne in mind

Other Statistics

No other continuous statistics of internal trade are published and in fact our knowledge of the subject is singularly defective

Trade statistics for foreign countries are compiled upon systems that are not altogether uniform See *London and Cambridge Economic Service Memorandum* No 21

CHAPTER XXVI

FINANCE

THE weekly *Bank of England Return* is published in the financial columns of the daily papers. Table 102 gives the version published by the *Times*.

TABLE 102

BANK OF ENGLAND RETURN FOR THE WEEK ENDED 28TH JULY, 1937

ISSUE DEPARTMENT

Notes issued—		Government Debt	£11,015,100
In circulation	£498,338,710	Other Government	
In banking department	28,067,915	securities	185,134,969
		Other securities	3,835,808
		Silver coin	14,123
		Fiduciary issue	200,000,000
		Gold coin and bullion	326,406,625
	£526,406,625		£526,406,625

BANKING DEPARTMENT

Capital	£14,553,000	Government securities	£114,410,022
Rest	3,551,532	Other securities—	
Public deposits[1]	10,528,723	Discounts and advances	5,811,909
Other deposits—		Securities	20,815,435
Bankers	104,259,233	Notes	28,067,915
Other accounts	37,322,336	Gold and silver coin	1,109,543
	£170,214,824		£170,214,824

	Amount	Increase or Decrease on Last Week	Increase or Decrease on Last Year
	£	£	£
Rest	3,551,532	+ 44,084	– 19,110
Public deposits	10,528,723	– 6,973,634	– 31,763,439
Other deposits—			
Bankers	104,259,233	+ 7,077,903	+ 29,036,162
Other accounts	37,322,336	– 827,401	– 2,049,135
Government securities	114,410,022	+ 6,061,325	+ 18,001,712
Other securities—			
Discounts and advances	5,811,909	– 18,177	– 1,326,393
Securities	20,815,435	– 2,553,606	+ 1,729,186
Reserve	29,177,458	– 5,068,590	– 23,200,027
Note circulation	498,338,710	+ 5,205,349	+ 49,767,951
Coin and bullion	327,516,168	+ 136,759	+ 86,567,924
Proportion	$19\frac{3}{16}$%	– $3\frac{4}{16}$	– $14\frac{5}{16}$

[1] Including Exchequer, Savings Banks, Commissioners of National Debt, and Dividend Accounts

The gold reserve (£328 mill.) is valued at the standard price of 84s 11½d per fine ounce. Revalued at the current price level (say 140s.) it would be worth £540 mill., an amount sufficient to redeem the whole of the note circulation. The fiduciary issue was reduced from £260 mill. to £200 mill. on 16th December, 1936, in order to offset a large gold purchase from the Exchange Equalization Fund. The Banking Department treats its stock of notes as equivalent to gold and on this basis the "Proportion" (of banking reserves to outside liabilities) is calculated as follows—

	28th July 1937	29th July 1936
	£000's	£000's
Outside Liabilities—		
Public Deposits	10 529	42 292
Bankers' Deposits	104 259	75 223
Other Accounts	37 322	39 371
Total (A)	15[illegible] 110	156 886
Banking Reserves—		
Notes	28 068	51 370
Gold and Silver Coin	1 109	1 007
Total (B)	29 177	52 377
Proportion (100B/A)	19[illegible]%	33⅓%

The "Proportion" for 28th July, 1937, is unusually low, the lowest in fact since 4th January, 1933. The fall is due to the record note circulation which has been allowed to deplete the Banking reserves.

If we amalgamate the two returns and consider the position as a whole we have the following figures—

	28th July 1937	29th July 1936
	£000's	£000's
Outside Liabilities—		
Note Circulation	498 339	448 571
Deposits, etc.	152 110	156 886
Total (A)	650,449	605 457
Reserves of Gold Coin and Bullion		
Issue Department	326,407	239 941
Banking Department	1 109	1 007
Total (B)	327 516	240 948
Reserve Ratio ("so-called") — 100B/A	50[illegible]%	39[illegible]%

Upon this basis, the situation has improved, the reason being that the note circulation is better covered than it was a year ago.

London Clearing Banks' Returns.

Monthly returns for the London Clearing Banks are published in the Press. A conveniently condensed version is given in the *Bank of England Statistical Summary*. (See Fig. 13 in Chapter VII.) A study of the table shows that the Banks can create credit in response to public demand by acquiring additional assets, viz. by granting loans, discounting bills, or purchasing securities. Experience has shown that a cash reserve equivalent to 10 per cent of deposit liabilities is sufficient for normal requirements, and the Clearing Banks aim at a percentage slightly over this figure. In pursuing this policy, the Banks must observe a due proportion between investments and liquid assets. Since nearly half of the cash reserves is represented by balances with the *Bank of England*, the volume of which is controlled by the monetary authorities, it is evident that the credit system of the country can be kept under close control, provided the Banks adhere to existing conventions.

The *London and Cambridge Economic Service* publishes a similar series in more abbreviated form. This series is confined to nine banks and avoids the break in continuity due to the inclusion of the two additional ones.

Bankers' Clearing House Returns.

Weekly returns of clearings by the London and Provincial clearing houses are published in the Press. The crude figures do not afford a satisfactory index of the state of trade, because they include Stock Exchange transactions and other operations in the capital market involving huge sums of money but having no appreciable bearing upon current commercial transactions. The difficulty may be partially avoided by elimination of the "Town" transactions of the London clearing house. The *Bank of England Statistical Summary* gives a monthly table showing total clearings (Metropolitan, Country and Provincial, (*a*) in Sterling and (*b*) in the form of an index number with and without seasonal adjustment).

The total turnover on customers' accounts is also published for the London Clearing Banks. In June, 1937, the figure amounted to £5,319 mill.[1]

[1] *Bank of England Statistical Summary*, July, 1937, p. 80

Other Financial Figures

Monthly returns of new capital issues are compiled by the *Midland Bank*. These figures are useful as a guide to market sentiment but throw little light on the amount of current investment in capital goods. A large variety of other financial information is available for which reference should be made to the *Bank of England Statistical Summary*, the *Federal Reserve Bulletin*, and the publications of the *League of Nations*.

Stock Exchange Securities.

Particulars of quotations and business done on the *London Stock Exchange* are published in the daily Press. Fuller information may be obtained from the London Official and Supplementary Lists, brokers' lists and the financial press. In addition there are subscription services which keep their clients posted with latest information as to dividends, yields, highest and lowest markings, etc.

Two problems present themselves in this connection—

1. Measurement of general level of Stock Exchange prices and levels in particular branches.
2. Measurement of volume of business done.

Measurement of Price Levels

The measurement of price levels is conveniently effected by the method of index numbers. No special difficulty attends the prices themselves, for the indices are generally confined to securities with a comparatively free market, but there are considerable complications as regards weights. Should weighting be based upon the amount of share capital involved, the importance of the industry as represented by its net output, or the number of shares traded upon the Stock Exchange? In the latter case should regard be paid to the normal amount of business done, or to the ebbs and flows of interest in particular markets and shares? Should the list of securities be fixed for a long period, or should it be revised frequently in order to reflect public sentiment?

It is impossible to generalize. One and the same system cannot apply to the investor who sticks to his holdings, the speculative investor who changes his holdings from time to time on long term views, and the speculator with short term views who aims at a quick

profit. The whole matter is at present in the experimental stage, and a variety of indices are forthcoming to suit various classes of operators.

The Actuaries' Investment Index.

This index has been compiled with great care and thought and is generally accepted as standard. Figures are supplied to subscribers only through a service conducted by the *Institute of Actuaries and Faculty of Actuaries* jointly, but summaries are published in the *Economist*. The securities comprised in the index are cross-classified according to type and status. Altogether there are 33 groups and 409 securities. The list is revised every January according to rules laid down beforehand. The minimum number of securities used to form a group is 3. Group price indices are calculated by an unweighted geometric average of price ratios. Thus if $p'_0\ p''_0\ p'''_0$ . . . represent prices at the base date and $p'_1\ p''_1\ p'''_1$ . . . prices at the current date and the number of securities is n, the value of the group index is proportional to

$$\left\{\frac{p'_1}{p'_0} \times \frac{p''_1}{p''_0} \times \frac{p'''_1}{p'''_0} \times \cdot \quad \cdot \quad \cdot\right\}^{\frac{1}{n}}$$

Combined price indices are formed by treating in the foregoing manner all the individual prices in every group to be included. The importance of each *group* as represented by the number of securities it has provided *is thus given weight in forming the combined* price index. Statistics of yield are the unweighted averages of the gross yields of the securities in each group. The base date is 31st December, 1928. Monthly figures are supplied for all fixed interest stocks and bonds and weekly figures for ordinary stocks and shares. Prices are the middle prices quoted in the *London Official List* on Tuesdays, adjusted to allow for accrued interest *less tax* up to the date of calculation. For ordinary shares the index of yield is based on earnings (as distinct from dividends) and adjustments are made when necessary in the published accounts in order to arrive at companies' normal earning capacities.

Other indices include—

The London and Cambridge Economic Service Index
The Bankers' Magazine Index.

The Investors Chronicle Index
The Financial News Index
The Financial Times Index

Each of these indices has its own special features and reference should be made to the publications concerned

Measurement of Business Done

The only available measure is the number of markings recorded in the daily Stock Exchange Lists

The number of markings affords a rough index of business but not a reliable one because—

1 Not all transactions are marked
2 Transactions vary in size
3 Several transactions may be included under one marking
4 Opportunities for marking depend upon pressure of business

CHAPTER XXVII

PRODUCTION

STRICTLY speaking, Production includes all activities associated with the manufacture and movement of goods until they pass into the hands of the final consumer. For Statistical purposes, however, it is usual to limit the term Production to the extractive and manufacturing industries. This distinction is based upon convenience rather than logic. The processes of extraction and manufacture yield tangible goods which can be readily counted or measured, whereas transport and distribution yield services which are sometimes embodied in tangible goods and at other times are rendered direct to consumers. In neither case are these services easy to identify.

Ideally, statistics of production (using the term in the narrow sense alluded to above) should refer to finished goods and to those alone. If production is measured at any intermediate stage there will be duplication owing to the subsequent reappearance of the intermediate goods in a finished form. In practice, however, this ideal is not attained, for three reasons—

1. The less advanced the state of manufacture the easier, as a rule, it is to collect statistics.

2. Statistics of intermediate goods are often more significant than statistics of final goods. A statement of the total amount of coal raised would include coal used in manufacture, including gas and electricity supply. But this statement would be intrinsically more interesting than a statement of the amount supplied to private consumers.

3. It is often difficult to identify consumers' goods. For instance, agricultural products are consumers' goods when consumed by human beings, and producers' goods when consumed by cattle. A motor-car frequently belongs to the category of producers' goods during working hours, and to that of consumers' goods during leisure.

In these circumstances it is usual to measure production at points found to be most convenient and where statistics are most easily

obta nable Any duplication inherent in this method will be corrected by devices to be described later

Methods of Measuring Production

There are two methods of measuring production

1 Periodical Censuses of Production aiming at a complete enumeration of the country s productive activities during a given period (generally a year) by quantity and value

2 Indices of Production aiming at approximate measurement of movements in production when census figures are not available Indices of production are usually confined to a relatively small number of items representing goods in a less advanced stage of manufacture and it is assumed that movements revealed by the indices reflect the movements of productive activities in general with sufficient accuracy

Census of Production

The *Board of Trade* has power to require statistics of manufacturing production under two Acts of Parliament viz the Census of Production Act 1906 and the Import Duties Act 1932 section 9 The former Act applies to all productive industry except Agriculture whilst section 9 of the latter Act only applies to factories or workshops making goods of a dutiable class or description The Board s powers under the two Acts differ somewhat as regards both matters for investigation and particulars enforceable This has led to complications and it is proposed to introduce further legislation with a view to assimilation of the two Acts and extension of the list of enforceable particulars Censuses of Production were held for the years 1907 1912 (completion prevented by the War) 1924 and 1930 [1] and Import Duty Act Inquiries for the years 1933 and 1934 [2] whilst for the year 1935 inquiries have been made under both Acts applying the Import Duties Act to all firms concerned with the kinds of goods to which that Act applied and the Census of Production Act to the remainder The results of the two inquiries will be consolidated into one report and any gaps in information caused by the fact that the powers under the Acts are not co extensive will be filled up as far as possible on a voluntary basis

[1] Fourth Census of Production 1930 Final Reports
[2] Reports on the Import Duties Act Inq iry 1933 and 1934

The Census covers production in the narrow sense, i.e. extractive and manufacturing operations, including building and the manufacturing activities of Government Departments. Transport, commerce, and professional and personal services lie outside its scope, whilst Agriculture, etc., forms the subject of a separate inquiry.

Census of 1935.

It is the *Board of Trade's* practice to issue the preliminary results of the Census in a series of supplements to the *Board of Trade Journal* at fortnightly intervals,[1] the object being to publish as quickly as possible those statistics which are considered to be of more immediate practical interest, and to follow them up with a series of final reports in volume form, designed on more elaborate lines. In order to indicate the scope of the present arrangements, it is proposed to give—

(1) A summary of the introductory notes to the Census of 1935.

(2) A specimen set of tables from the preliminary reports relating to the iron and steel trade (blast furnaces).

(3) Further particulars based on the 1930 Census, covering matters that have not yet appeared in the 1935 Census.

Summary of Introductory Notes (1935).

Scope of the Census. The Census covered manufacturing industries, mines and quarries, the building and contracting trades and the productive services carried out by public utility undertakings, whether publicly or privately owned. Repairing and processing work was included. The number of separate trades distinguished was 122. All the particulars relate to the United Kingdom.

Exclusion of Small Firms. In order to expedite the publication of the results of the Census and to reduce its cost, firms employing not more than ten persons as a yearly average were exempted from the obligation to make detailed returns, the only information required being a statement of the nature of their business and the average number of their employees during the year.

Comparisons with Previous Years. Comparison of the results for 1935 is made with those of the most recent previous year for which statistics are available, i.e. with those for the year 1934 in the case

[1] The 1935 series began in the issue for 28th January, 1937.

A novel feature of the 1935 Census is the inclusion of details of important materials classified by quantity and value For previous years the cost of materials was stated in a fump sum and details are not usually available

(viii) *Work Given Out* Firms giving out work were required to state the aggregate amount paid on that account during the year The amount returned was not to include the cost of any of the items accounted for under the heading of materials

(ix) *Net Output* The Net Output of a trade is the figure which results from deducting from the value of the gross output the aggregate of the cost of materials used and amount paid for work done together with the amount of any Excise duty included in the value of the products and not in that of the materials used This figure represents the value added to materials by the industrial processes and after allowance for a sum sufficient to cover the depreciation of plant and machinery constitutes the fund from which wages salaries rents royalties rates and taxes advertisement and selling expenses and all similar charges have to be provided as well as profits

(x) *Persons Employed* The general nature of the returns will be evident from the specimen table

xi) *Outworkers* In those trades employing outworkers particulars were required of the number of outworkers employed at two dates covered by the return

(xii) *List of Trades Covered* This is too lengthy to quote and reference must be made to the original reports

Mechanical Power The 1930 reports gave particulars of mechanical power used under the headings of (a) prime movers (b) electric generators and (c) electric motors No similar particulars were collected in 1935 but on the other hand additional particulars have been obtained with regard to fuel consumption

TABLE 103

Census of Production (U.K.) of 1935

Iron and Steel Trades

THE IRON AND STEEL TRADE (BLAST FURNACES)[1]

TABLE I
General Summary

Particulars	Unit	1935	1934
Value of products (gross output) . .	£'000	21,047	18,540
Cost of materials, fuel, and electricity used	,,	16,964	14,949
Net output	,,	4,083	3,591
Average number of persons employed .	No.	15,815	14,875
Net output per person employed . .	£	258	241

TABLE II
Output of Principal Products

Kind of Goods	1935		1934	
	Quantity	Value	Quantity	Value
	Th. Tons	£'000	Th. Tons	£'000
Pig iron—				
Forge	115·4	345	122·4	339
Foundry.	1,320·4	4,097	1,331·0	3,957
Acid (hematite) . .	1,523·8	4,840	1,476·0	4,572
Basic	3,412·1	9,117	2,836·6	7,497
Total—Pig iron . .	6,371·7	18,399	5,766·0	16,365
Ferro-alloys—				
Ferro-manganese and spiegeleisen . . .	95·5	797	81·2	716
Ferro-tungsten . .	2·2	552	15·4	743
Ferro-molybdenum . .	0·8	263		
Other[2]	2·5	299		
Total—Ferro-alloys .	101·0	1,911	96·6	1,459
Total—Principal products . . .	6,472·7	20,310	5,862·6	17,824

[1] *Board of Trade Journal*, 11th March, 1937.
[2] Including ferro-silicon, ferro-titanium, and ferro-vanadium.

TABLE III

PRODUCTION EXPORTS AND IMPORTS

Kind of Goods		Production	Exports	Retained Imports
Pig iron—		Th Tons	Th Tons	Th Tons
Forge	1935	115 4	2 2	[1]
	1934	*122 4*	*1 3*	—
Foundry	1935	1 320 4	90 5	46 3[2]
	1934	*1 331 0*	*73 2*	*46 1[2]*
Acid (hematite)	1935	1 523 8	48 8	—
	1934	*1 476 0*	*47 9*	—
Basic	1935	3 412 1	0 8	37 5
	1934	*2 836 6*	*0 8*	*79 4*
Ferro-alloys—				
Ferro manganese and spiegeleisen	1935	95 5	11 7	2 8
	1934	*81 2*	*9 0*	*4 4*
Other kinds	1935	5 5	2 9	40 8
	1934	*15 4*	*0 9*	*34 5*

The following tables relate only to firms whose returns were made on schedules for Blast Furnaces and are summarized in Table I

TABLE IV

OTHER OUTPUT OF THE IRON AND STEEL TRADE (BLAST FURNACES)

The value of the gross output recorded on schedules for Blast Furnaces was £21 047 000 in 1935 and £18 540 000 in 1934 of which £19 210 000 in 1935 and £17 152 000 in 1934 consisted of products included in Table II Particulars of the remaining items are shown in the following table

Kind of Output	1935		1934	
	Quantity	Value	Quantity	Value
	Th Tons	£ 000	Th Tons	£000
Iron castings in the rough	5 0	22	3 3	13
Ground slag	428 8	112	522 1	131
Tarred macadam	180 7	98	281 6	158
	Mill B T U (Kw hrs)		Mill B T U (Kw hrs)	
Electricity sold	378	607	210	314
Gas sold	—	740	—	624
Other output	—	258	—	138
Total	—	1 837	—	1 388

[1] Less than 50 tons

[2] Including Pig iron smelted wholly with charcoal (11 900 tons in 1935 12 000 tons in 1934) and Vanadium titanium pig iron produced in an electric furnace (4 800 tons in 1935 3 400 tons in 1934)

TABLE V

MATERIALS, FUEL, AND ELECTRICITY PURCHASED AND USED

Kind of Materials, etc.	1935 Quantity	1935 Cost	1934 Cost
	Th. Tons	£'000	£'000
Materials used—			
Iron ore	14,982·1	7,415	
Limestone	1,745·5	492	
Cinder and scale	474·4	330	
Purple ore	252·9	198	
Scrap	276·0	462	
Fuel and electricity for all purposes—			14,949
Coal	344·0	220	
Coke	7,170·3	6,368	
	Th. B.T.U. (Kw.-hrs)		
Purchased electricity[1]	128,821	185	
Other purchased materials and fuel	—	1,294	
Total	—	16,964	14,949

TABLE VI

CONSUMPTION OF ELECTRICITY IN 1935

Electricity Consumed	1935
	Th. B.T.U. (Kw.-hrs)
Generated in same works	190,119
Generated in other works under same ownership	71,349
Purchased	57,472
Total	318,940

[1] Including electricity generated in other works under the same ownership.

meet with success The same remark applies to statistics of Capital employed

General Results—Census of 1930.

As the general results of the 1935 Census are not available at the time of writing, it is necessary to substitute those for 1930 A condensed version is shown in Table 104

Other Production Statistics.

Statistics regarding the production of the principal commodities are obtainable from a variety of sources, both for this country and for foreign countries They do not call for special comment [1]

Indices of Production.

The function of an Index of Production is to bridge over the gaps between successive Censuses If we collect representative figures of production from the principal industries convert them into index numbers and weight the latter on the basis of the results disclosed at the last census of production, the results will measure changes in industrial activities with sufficient accuracy to enable them to be used with confidence over a short period, while they can always be trued up as the successive Censuses appear

The pioneer work in this direction was performed by the *London and Cambridge Economic Service* which has published an Index of Production continuously since 1924 In 1928 the *Board of Trade* began publication of a new and independent index, which is now accepted as standard for general statistical purposes The present account is confined to the *Board of Trade* Index

Board of Trade Index of Industrial Production.

The following is an analysis[2] *of the factors taken into account* when the *Board of Trade Index* was designed

The objective of study is the *net output* of the various industries, i e the excess of the value of the products over the value of materials used up in their manufacture The determination of the *net output*

[1] The best summary is to be found in the tables illustrating the economic position of the United Kingdom published monthly in the *Board of Trade Journal*

[2] See Flux Indices of Productive Activity " *J R S S* Vol xc (1927) pp 225–258

TABLE 104. INDUSTRIAL PRODUCTION

General Results of the Censuses of Production of 1924 and 1930 Classified in Principal Industry Groups[1]

[*Compiled from the Final Report on the Fourth Census of Production* (1930)]

Trade Group (1)	Year	Gross Output (Selling Value of Goods Made and Value of Work Done) (2)	Cost of Materials and Fuel Used and Amount Paid for Work Given Out (3)	Net Output (Excess of Col. 2 over Col. 3)[2] (4)	Average Number of Persons Employed (except Out-workers) (5)	Net Output per Person Employed (6)	Power Available: Prime Movers (7)	Power Available: Electric Motors Driven by Purchased Electricity (8)
		£'000	£'000	£'000	No.	£	Th. H.P.	Th. H.P.
Factory Trades—								
Iron and Steel	1924	295,445	196,801	98,644	498,912	198	2,183·4	711·9
	1930	237,695	145,967	91,728	493,577	186	2,169·9	981·3
Engineering, Shipbuilding and Vehicles	1924	402,155	203,749	198,406	985,578	201	467·3	1,257·8
	1930	461,331	231,346	229,985	1,074,749	214	287·1	1,638·0
Non-Ferrous Metals	1924	92,333	67,061	25,272	114,988	220	120·9	141·1
	1930	107,590	83,994	23,596	109,718	215	116·3	209·7
Textiles	1924	762,826	541,030	221,796	1,261,984	176	2,347·2	404·0
	1930	432,387	283,385	147,402	1,062,250	139	2,149·2	623·2
Leather	1924	42,070	30,441	11,629	48,429	240	37·0	35·4
	1930	36,017	25,837	10,180	46,146	221	34·9	43·3
Clothing	1924	183,227	107,508	75,719	473,968	160	47·0	66·0
	1930	180,499	102,352	78,147	492,124	159	36·0	87·7
Food, Drink and Tobacco	1924	669,532	406,908	172,454	439,787	392	317·1	282·5
	1930	663,947[3]	403,222	188,205[3]	472,437	398	324·2	421·1
Chemicals, etc.	1924	194,062	125,078	65,805	178,094	369	342·0	161·2
	1930	181,752	103,621	72,921	178,151	409	555·4	296·9
Paper, Printing and Stationery	1924	161,603	67,644	93,884	342,649	274	295·1	181·3
	1930	177,373	73,984	103,309	380,003	272	484·4	294·2
Timber	1924	59,387	32,055	27,332	137,554	199	115·6	99·1
	1930	68,708	37,243	31,465	167,812	188	109·7	170·1
Clay and Building Materials	1924	68,900	25,330	43,570	208,948	209	300·0	103·0
	1930	71,800	26,699	45,101	224,516	201	319·1	256·6
Miscellaneous	1924	94,474	53,023	41,451	165,746	250	200·4	130·7
	1930	91,868	48,950	42,918	174,076	247	183·3	225·6
Total— Factory trades	1924	3,026,014	1,856,628	1,075,962	4,856,637	222	6,773·0	3,574·0
	1930	2,710,967	1,566,600	1,064,957	4,875,559	218	6,869·5	5,247·7
Other Trades—								
Building and Contracting	1924	162,725	82,131	80,594	419,053	192	89·3	87·6
	1930	194,288	100,223	94,065	453,807	207	97·5	125·8
Mines and Quarries	1924	273,037	46,634	226,403	1,280,984	177	3,310·7	522·6
	1930	187,314	32,140	155,204	1,018,844	152	3,274·1	639·2
Public Utility Services and Government Departments	1924	285,727	120,056	165,671	741,441	223	6,500·9	166·4
	1930	313,483	122,463	191,020	793,224	241	10,968·6	294·8
Total—Other Trades	1924	721,489	248,821	472,668	2,441,478	194	9,900·9	776·6
	1930	695,115	254,826	440,289	2,265,875	194	14,340·2	1,059·8
Total— All Trades	1924	3,747,503	2,105,449	1,548,630	7,298,115	212	16,673·9	4,350·6
	1930	3,406,082	1,821,426	1,505,246	7,141,434	211	21,209·7	6,307·5
England and Wales	1924	3,313,955	1,863,449	1,363,793	6,361,905	214	145,405·7	3,612·0
	1930	3,022,349	1,606,880	1,341,667	6,278,036	214	18,670·0	5,408·5
Scotland	1924	366,124	198,522	160,891	779,218	206	1,940·2	697·2
	1930	324,380	175,606	143,166	717,262	200	2,276·5	822·8
Northern Ireland	1924	67,424	43,478	23,946	156,992	153	228·0	41·4
	1930	59,353	38,940	20,413	146,136	140	263·2	76·2

[1] The figures relate to firms and undertakings that employed more than ten persons on the average except in the case of Northern Ireland, where they cover firms employing more than five persons.
[2] Excluding estimated Excise duty.
[3] Including subsidy on home-grown sugar.

Source: Statistical Abstract 1937 (Cmd. 5353).

requires information not readily obtainable, particularly for short periods of time, whilst figures of gross output demand relatively little analysis of the records of business turnover. The gross output of any particular industry may be expected to bear to the net output a relation which changes by relatively slow degrees except in cases in which considerable and rapid changes in structure and organization are taking place.

Should it be found by experience that the changes that take place are of slow development, the intervals which may be permitted to take place between successive general inquiries [sc. Censuses of Production] may be extended without entailing a serious sacrifice of usefulness in the information secured; but if the changes are *found to proceed rapidly, it may be desirable to shorten the intervals.*

The well known principle that very great precision in weighting is not essential may be recalled, and the factors of relative importance of the different industries, which serve for the purpose of combining these individual index numbers to form a general index of industrial activity, remain useful, even though changes of moderate amount may have taken place in the industrial distribution of productive energies.

A point of at present unknown importance arises, however, in considering the changes which are apt to occur in the course of the cycle. At the period of greatest activity, industries concerned with capital goods undergo exceptional expansion, whilst in times of depression it is these industries which experience with greatest sharpness restriction of demand.

The relation of gross output to net output is also likely to be affected by the same class of change, and an index of activity based on gross output may fall off less in times of depression and rise less *in times of prosperity than corresponds to the real changes in* wealth production.

In practice, it may be found not only that data of net output are lacking, but that data of gross output are available only to a limited extent, and in such cases it may be desirable to make use of indirect indications of the ebb and flow of activity. This leads to a discussion of the possibilities of measurement by materials utilized, machine activity, power employed, workers employed, orders received, and sales or deliveries.

Agriculture is excluded from the inquiry for several reasons

The fluctuations of agricultural production are in so great a degree dependent on seasonal climatic conditions that the yields, so far as quantity is concerned, do not bear any close relation to the variations in effort expended on them year by year; and in any case the figures of production are suited rather to an annual inquiry than to one of greater frequency.

An effort should be made to include as wide a range of industries as possible, the aim being to secure an adequate representation of the industry actually carried on, subject to the exclusion of industries of trifling importance. It is not essential that the range of industries should be identical in each of the countries between which it is desired to make comparisons. The index for each country should be as fully representative as possible of the conditions of that country, and if that end is secured, the largest degree of comparability will be attained.

Details of Board of Trade Index.

This index was begun in 1928, with base year 1924, and the original series ran quarterly from 1928 to 1934 inclusive. In 1935 the index was revised in the light of information available from the 1930 Census of production as to the reliability of the existing series, this revision comprising (1) a review of the data for the trades or sections of trades covered by the existing index, and (2) the use of new series representing output not previously included within the scope of the index, the principal addition being that in respect of the construction of new buildings. The following account relates to the index in its present form.

The method adopted has been to compare the best available figures measuring the volume of production in each industry in each year and in each quarter with the corresponding figures for the whole year 1930. The indices of activity thus obtained for the different branches of trade have been combined so as to form indices of activity for the leading groups of industries, and for industry as a whole. The principle adopted in combining the individual indices has been to assign to each its relative importance as measured by net output (i.e. the value added in production or manufacture to the materials used) as ascertained at the 1930 Census of Production. The indices may thus be said to have represented the variations in the volume of net output as compared with 1930.

The information from which the indices have been constructed has been obtained from voluntary returns furnished by trade associations and by individual firms from official returns of imports and exports of wages paid and production from the bulletins of certain industrial federations and from trade papers in which production and movements in stocks are shown

The various industries for which information has been obtained have been classified into groups which are comparable as far as possible with the grouping adopted for the 1930 Census of Production The groups are as follows—

(1) Mines and Quarries (2) Iron and Steel (3) Non ferrous Metals (4) Engineering and Shipbuilding (5) Building Materials and Building (6) Textiles (7) Chemicals Oils etc (8) Leather and Boots and Shoes (9) Food Drink and Tobacco (10) Gas and Electricity In addition to the industries enumerated above particulars of the production of pianos and paper and of the consumption of rubber are included in the calculation of the general index

The sections of industry covered by the information at present received represent over 70 per cent of the total manufacturing and mining activity of the United Kingdom and apart from building about 90 per cent of the total activity of the groups of industry set out above For building the proportion is about 30 per cent Of the branches of trade not covered by the data summarized in the table the most important are the clothing trade (other than boots and shoes) and public utility services other than gas and electricity

Table 105 shows details for the years 1935 and 1936 and for each of the five quarters ending March 1937[1]

Comments on the Board of Trade Index

The reliability of the index depends upon the proposition that movements of a random sample of data will tend to follow the movements of the complex from which it is drawn the necessary element of randomness being supplied by the accidents which make certain data available As the actual behaviour of the index fits in with information available from other sources as well as with facts of general observation this claim appears to be justified and we may conclude the index measures what it purports to measure As the

[1] *B T J* 20th May 1937 (p 700)

TABLE 105

BOARD OF TRADE INDEX OF PRODUCTION—
UNITED KINGDOM (1930 = 100)[1]

Group	Year 1935	Year 1936	1936 March Quarter	1936 June Quarter	1936 Sept. Quarter	1936 Dec. Quarter	1937 March Quarter
1. Mines and quarries	91·7	94·4	100·6	88·4	89·7	99·1	99·5
2. Iron and steel	125·6	150·1	146·2	149·5	149·1	155·6	158·0
3. Non-ferrous metals	137·3	143·8	134·8	140·9	145·3	154·0	154·2
4. Engineering & shipbuilding	104·8	123·1	116·3	122·4	121·6	132·3	136·1
5. Building materials and building	147·0	157·1	148·8	157·8	164·7	157·3	147·7
6. Textiles	119·1	126·4	127·3	124·9	123·3	130·3	130·5
7. Chemicals, oils, etc.	110·6	114·0	115·1	111·7	110·2	119·1	120·5
8. Leather and boots and shoes	116·0	120·7	126·1	121·0	116·4	119·6	120·6
9. Food, drink and tobacco	107·6	114·5	106·0	114·5	115·2	121·2	113·3
10. Gas and electricity	132·6	148·2	(a)	(a)	(a)	(a)	(a)
Total of manufacturing industries (2–10)[2]	117·0	129·4	126·9	129·2	127·8	137·6	137·3
Total of all groups (1–10)[2]	113·5	124·6	123·2	123·4	122·4	132·1	131·9

data are based for the most part on physical output and sterling values (when they occur) are corrected for price movements, it is clear that the index relates to physical volume and throws no light (at least directly) upon the profitability of the activities concerned. It should be emphasized that the index only covers production in the narrow sense, i.e. Distribution and Transport are excluded. Methods of constructing a general index of business activity in which these items are included are discussed in Chapter XXIX.

Course of the Board of Trade Indices of Production.

The following table shows the course of the indices over the period 1928–36. In order to obtain a continuous series, the old series (base year 1924) has been converted to base year 1930 by application of the constant factor 0·9690 (= the reciprocal of 1·302). The converted figures are shown in *italics*.

[1] *Board of Trade Journal*, 20th May, 1937 (p. 700).
[2] Includes also various industries not specified above.
(a) Quarterly particulars of gas production are not available and complete information in respect of the year 1936 cannot yet be given; a provisional estimate of the quantity made has been used for the calculation of the group index for the year 1936. The available data for electricity have been included in the general index numbers for the four quarters of 1936 and the first quarter of 1937.

TABLE 106

UNITED KINGDOM—COURSE OF BOARD OF TRADE INDEX OF INDUSTRIAL PRODUCTION

(All Groups)

Year and Quarter (1)		Old Series *(1924 = 100)* (2)		*Converted* or New Series *(1930 = 100)* (3)	
1928	I	109·3	105·5	*105·9*	*102·2*
	II	103·6		*100·4*	
	III	100·2		*97·1*	
	IV	108·4		*105·0*	
1929	I	110·6	111·8	*107·2*	*108·3*
	II	112·0		*108·5*	
	III	110·7		*107·3*	
	IV	114·0		*110·5*	
1930	I	111·0	103·2	*107·6*	*100·0*
	II	103·1		*99·9*	
	III	99·5		*96·4*	
	IV	99·0		*95·9*	
1931	I	94·6	*93·7*	*91·7*	*90·8*
	II	92·1		*89·2*	
	III	89·3		*86·5*	
	IV	97·3		*94·3*	
1932	I	95·0	93·3	*92·1*	*90·4*
	II	94·3		*91·4*	
	III	87·4		*84·7*	
	IV	95·0		*92·1*	
1933	I	94·8	98·6	*91·9*	*95·5*
	II	96·7		*93·7*	
	III	96·8		*93·8*	
	IV	105·0		*101·7*	
1934	I	110·3	110·8	105·7	106·1
	II	110·3		104·6	
	III	106·0		103·2	
	IV	116·0		111·9	
1935	I			113·0	113·5
	II			111·5	
	III			110·7	
	IV			120·7	
1936	I			123·2	124·6
	II			123·4	
	III			122·4	
	IV			132·1	
1937	I			131·9	

International Comparisons.

Most industrial countries publish indices of production. As these are compiled on different systems and cover different varieties of goods, comparisons are difficult and can only be made on broad lines. The most convenient sources of information are to be found in the *League of Nations* publications.[1]

[1] *Monthly Bulletin of Statistics; Statistical Year-book; World Production and Prices.*

CHAPTER XXVIII

WEALTH

National Income[1]

The National Income is the estimated value of the flow of goods consumed by, and services rendered to, residents in the United Kingdom during the year, plus (or minus) additions to (or subtractions from) the National Capital or Stock of material goods[2]

To avoid complications we confine the goods and services in question to those customarily exchangeable for money The main effect is to exclude unpaid domestic services of women Goods and services produced commercially are reckoned at selling values and others (mainly public services) at cost of production Goods are reckoned to be consumed when (having reached the finished stage) they pass into the hands of final consumers There is one exception private dwelling houses are invariably treated as capital goods even if owner-occupied and only the annual yield is brought into the calculation

National Income may be reckoned Gross or Net the difference representing allowances due for all forms of wastage of capital equipment Provision is made in the calculations for writing down stocks in periods of falling prices

Methods of Estimation

Three main methods are available—

(1) *Aggregation of Incomes* Since all goods and services brought into the calculation must *ex hypothesi* be paid for the total of consumption plus investment must be equivalent to the total of money incomes provided we exclude from the latter any individual incomes not corresponding to the creation of tangible services This rule excludes all gratuitous payments including public relief

[1] For a general discussion of the whole subject see Sir Josiah Stamp Methods Used in Different Countries for Estimating National Income *J R S S* Vol xcvii pp 423 66 and the continuation of the discussion in pp 541–57 Also Colin Clark s *National Income and Outlay* (1937)

[2] The orthodox definition is in terms of things produced This involves complications due to transactions with foreigners and the above is to be preferred

and private allowances. In addition it is usual to exclude war pensions and interest on non-productive national debt on the ground that the services to which they correspond were not rendered during the current year.

Personal incomes above the income tax exemption limit and most impersonal incomes may be estimated from the annual reports of the *Board of Inland Revenue*, and the National wage bill from material collected by the *Ministry of Labour*. The residue, consisting of small salaries and incomes of small traders, etc., is estimated in the best information available.

(2) *Consumption plus Investment.* By this method the total value of goods consumed and services rendered is estimated from a variety of sources, while the value of Investment (i.e. actual outlay on capital equipment, etc.) is estimated from the Census of Production.

(3) *From Output Statistics.* Here the basis is the Net Output as shown by the Census of Production. To this must be added estimates of values created by Agriculture, Transport, the Distributive Trades, the Professions and personal services, obtained from a variety of sources.

Recent estimates of the National Income of the United Kingdom have been made by Bowley and Stamp, Colin Clark, Coates and Flux. As is usual in cases of this kind, there are considerable differences between the various estimates, and detailed reconciliation is out of the question.

A useful series of estimates is that given by Colin Clark. (See Table 107.)

Distribution of Income.

Statistics of the distribution of incomes amongst individuals are available for persons liable to sur-tax, i.e. persons with incomes of £2,000 a year and upwards.[1]

Statistics of the distribution of incomes from £130 a year upwards were prepared by the Board of Inland Revenue for the fiscal years 1918–19 and 1919–20,[2] but the publication of these tables has been discontinued on the ground of expense.

[1] See *Seventy-eighth Report of the Commissioners of His Majesty's Inland Revenue for the Year Ended 31st March*, 1935 (Cmd. 5015) (1936).

[2] See *Report of Committee on National Debt and Taxation*, (1927) Appendix XIV; and Stamp: *Wealth and Taxable Capacity*, p. 81.

TABLE 107

QUARTERLY FIGURES OF NATIONAL INCOME OF THE UNITED KINGDOM FREE FROM SEASONAL VARIATION

£ Mill per Quarter

Year and Quarter	Consumption Plus Investment Method	Output Statistics Method
1929—		
3	1 229	1 248
4	1 222	1 228
1930—		
1	1 181	1 195
2	1 148	1 171
3	1 170	1 147
4	1 141	1 117
1931—		
1	1 135	1 091
2	1 098	1 084
3	1 107	1 091
4	1 065	1 085
1932—		
1	1 067	1 074
2	1 056	1 065
3	1 053	1 056
4	1 051	1 063
1933—		
1	1 038	1 036
2	1 067	1 070
3	1 091	1 092
4	1 128	1 120
1934		
1	1 135	1 139
2	1 150	1 148
3	1 158	1 156
4	1 194	1 169
1935—		
1	1 217	1 196
2	1 238	1 227
3	1 250	1 225
4	1 286	1 259
1936—		
1	1 301	1 292
2	1 302	1 343

[1] Colin Clark *National Income and Outlay* (1937) p 206. This series is continued in a somewhat different form in Pritchard Wood's *Commercial Barometer*

The Board however, publish estimates of the total number of incomes above the exemption limit.[1]

Estimates of the distribution of personal incomes in the United Kingdom for the years 1929 and 1932 have been made by Colin Clark. The distribution of incomes above the £2,000 limit is known precisely from the Sur-tax statistics. The total number of incomes over the exemption limit is estimated annually by the *Board of Inland Revenue*, and the number over £250 may be estimated from National Insurance statistics. Intermediate values are interpolated using a *Pareto* curve.[2]

[1] See *Seventy-eighth Report of the Commissioners of His Majesty's Inland Revenue for the Year ended* 31*st March*, 1935 (Cmd. 5015) (1936), pp 59 and 64.
[2] Colin Clark: *National Income and Outlay*, 1937, pp. 102–115

CHAPTER XXIX

BUSINESS BAROMETERS AND BUSINESS ACTIVITY INDICES

MUCH attention is now being given by economists and statisticians to the study of industrial fluctuations or to use a convenient continental term Conjuncture

Types of Investigation

According to Professor Jones[1] investigators may be divided into three groups

The first group to which most economists belong comprises those who are mainly interested in causes and assuming fluctuations to be an evil in seeking remedies The second group of investigators consists of those who are mainly concerned with the actual course of trade and the possibility of measuring fluctuations in business activity *The third group consists of those who are mainly concerned* not with the actual course of trade in the past but with the probable course of trade in the immediate future They are the business forecasters

These three groups are not mutually exclusive and for our present purposes it will not be necessary to exercise any special discrimination We shall be concerned with immediate statistical principles rather than with ultimate aims and designs

Methods of Analysis

There are now in existence a large number of statistical series giving continuous information more or less representative upon current economic phenomena and there are three ways in which these series can be employed to convey definite impressions

1 To tabulate or plot a large selection of representative series without attempting any form of combination This may be accompanied by a verbal summary and interpretation of the data or the reader may be left to draw his own conclusions

2 To select (say) three or four individual series supposed to be peculiarly sensitive to current economic events

[1] Jones Business Forecasting *E J* Vol 38 (1928) p 414 *seq*

3. To combine an assortment of these series into a single index upon the hypothesis that significant movements will accumulate in the result and that non-significant movements will cancel out.

The statistical treatment of these series varies according to circumstances. The data may be presented *simpliciter* or in the form of indices. They may be corrected by the elimination of trend (i.e. long term) movements, or for seasonal variations. Currency values may be deflated, i.e. adjusted to the pre-war purchasing power of money. There may be adjustments for the growth of population.

The literature of the subject is voluminous, and since no new statistical principles are involved, it will be sufficient to indicate the chief current sources of information.

London and Cambridge Economic Service.

This is a subscription service for the benefit of economists, business men, and others requiring a reasoned and authoritative account and interpretation of current economic movements. The service is administered by an executive committee of seven persons and an editorial committee of twelve persons drawn from the academic staffs of the University of Cambridge (Economics Department) and the London School of Economics. The service consists of the following—

(1) A monthly bulletin comprising: (*a*) the United Kingdom Index Chart; (*b*) the U.S.A. Index Chart; (*c*) a review of the general business position in the United Kingdom; (*d*) analyses of recent movements in the United Kingdom and U.S.A.; (*e*) a series of tables and graphs illustrating the course of Finance, Prices and Wages, Trade and Transport and Unemployment; and (*f*) sundry other features.

(2) A special quarterly issue of the Monthly Bulletin comprising in addition a review of Finance, Trade and Production during the quarter. In this issue the usual graphs and tables are set out in extended and elaborated form

(3) A supplement to the Monthly Bulletin, giving additional figures for the United Kingdom available since the last monthly issue, together with reviews of the position in foreign countries contributed by expert foreign correspondents.

(4) A special quarterly issue of the supplement setting out the information in more elaborate form.

(5) Special memoranda on topics of general economic interest

The series is available to subscribers only, but by special arrangement with the *Royal Economic Society*, Fellows of that society are supplied with off prints of certain bulletins a few weeks after publication of the originals

The tables are clearly laid out with full references to sources of information and especial care is taken to ensure continuity The running commentary is incisive and well informed, although perhaps too technical and too guarded in its terms to appeal to the general business community A specimen Index Chart is shown in Fig 44

INDEX CHART, U K.

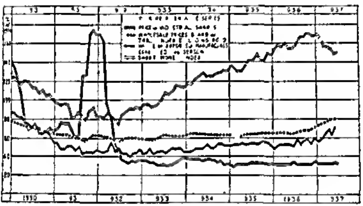


FIG 44

Board of Trade Tables Illustrating the Economic Position in the United Kingdom

Table I of this series comprises eleven main heads (with subdivisions) relating to matters of general economic interest Table II comprises thirteen main heads (with subdivisions) relating to particular trades The data are arranged in a form convenient for comparison and show (1) actual figures, (2) index numbers with base year 1930 Publication takes place monthly in the *Board of Trade Journal* These tables have been published continuously in the same form for nine years and provide a valuable source of information for those wishing to compile a continuous series of figures extending several years back Care must be taken on two

points. (1) Current figures are liable to revision in later issues. Substantial alterations are marked with an asterisk, but minor ones are effected without comment; (2) in the year 1935 the base year of the index numbers was changed from 1924 to 1930. There is risk of overlooking this fact if one is in a hurry.

Bank of England Statistical Summary.

This publication provides a concise and comprehensive monthly summary of current data in the fields of Banking and Finance, Prices, Employment, External Trade, Industrial Activity, Security Prices, Wages, etc. The tables are compiled from authoritative sources under expert supervision and great care has been given to the question of lay-out. Each month's issue includes special features based on important periodical returns from official and other sources. The numerous diagrams are neat, decisive, and intelligible. This is an admirable publication from all standpoints, and the student would do well to take it as his model.

Ministry of Labour Charts Illustrating the Course of Trade, etc.

The *Ministry of Labour* publishes a quarterly series of Charts illustrating the course of Trade, Output, Prices, Wages Finance and Unemployment, together with tables giving supporting figures. The data are shown in the form of quarterly averages from the year 1928 onwards. Index numbers are converted to base = 1924. The results are clear and compact. The adoption of the quarterly basis eliminates a certain amount of troublesome fluctuation and brings out the trends but involves the drawback that the figures are continually behindhand

"Trends."

"Trends" is a monthly graphical review of business movements. Originally established by Messrs. Harold Whitehead and Staff, Business Consultants, it is now incorporated with *Industry Illustrated*. The service consists of a number of graphs of economic series with short commentaries. The general style and lay-out are attractive, but the absence of numerical data detracts from the value of the service

The "Economist" Index of Business Activity[1]

Publication of this index began in October 1933 with base year = 1924 In July 1936 it was revised and recalculated with base year = 1935 The object of the index is to measure changes in the economic activity of the United Kingdom in quantitative—not monetary—units in other words it is designed to give an approximate idea of fluctuations in the real national income

With three exceptions all the series are on a daily average basis—thus eliminating the effect of months of varying length The exceptions are employment where the figures relate to a particular day motor vehicles where the index represents the number of licences current during the month and building activity where a twelve months moving average is employed All the series except building (where the correction would be superfluous) are corrected for seasonal fluctuations the seasonal correctors being obtained by averaging the percentage deviations of the monthly figures from a twelve months moving average over a period of about ten years These seasonal coefficients are recalculated each year to take into account the fluctuations of the previous year The final index is a weighted geometric average of the constituent series The choice of a geometric mean while giving a higher degree of accuracy over a period of time makes the figures for 1926 somewhat arbitrary During the general strike period two of the indices fell to zero which on a strict computation brings the final index down to zero also The compilers have therefore adopted the expedient of ignoring these two series and of basing the index for these months on the remaining series The figures for the whole of 1926 should however be used with reserve

The weights allotted to particular series were determined by the rough balancing of four main considerations the importance of the sphere of activity represented by the series its excellence as a measure of general business activity its degree of freedom from sudden and arbitrary movements and its statistical accuracy None of these considerations can be accurately assessed The final allotment of weights is therefore necessarily arbitrary but it is believed that the following weighting now in use in the index is not inherently unreasonable employment 10 coal 4 electricity 2 merchandise on railways 4 commercial motors 2 postal

[1] See the *Economist Trade Supplement* 25th July 1936 (No 158)

receipts, 3; building activity, 2; iron and steel, 2; cotton, 1; imports of raw materials, 2; exports of British manufactures, 3; shipping movements, 2; metropolitan, country and provincial bank clearings, 4; town clearings, 1.

The component series are calculated as follows—

1. *Employment.* Based on the Ministry of Labour estimates of insured workers aged 16 to 64 in employment in Great Britain. Seasonal fluctuations are eliminated.

2. *Consumption of Coal.* From the output of saleable coal is deducted exports of coal, coke, and manufactured fuel, and of coal shipped for the use of steamers. The resulting figure of home consumption is adjusted for variations in stocks of coal at pithead and is placed on a daily basis. Seasonal fluctuations are eliminated.

3. *Industrial Consumption of Electricity.* From the daily average output of the "authorized undertakers" is deducted the units generated for domestic purposes and for use in traction and public lighting. The series is adjusted for transfer from private to public generation and is corrected for seasonal fluctuations.

4. *Merchandise on Railways.* Based on a daily average of the tonnage of freight (less coal and coke) carried on the standard-gauge railways of Great Britain. Seasonal fluctuations are eliminated.

5. *Commercial Motor Vehicles in Use.* The monthly computations of the number of licences for goods vehicles current in Great Britain, published by the Society of Motor Manufacturers and Traders, are corrected for seasonal fluctuations.

6. *Postal Receipts.* The official series of daily postal receipts, calculated by the Post Office, is corrected for seasonal variations.

7. *Building Activity.* From the Ministry of Labour figures showing the value of building plans approved by 146 local authorities in Great Britain a twelve-months' moving average is calculated, on the assumption that about twelve months elapse, on the average, between the passing of the plans and the completion of the building. Seasonal fluctuations are automatically eliminated by this process, but no account can be taken of the fact that some plans approved may not be executed. The series is corrected for changes in building costs by means of an index, the construction of which was explained in the *Economist* of 11th November, 1933.

8. *Consumption of Iron and Steel.* Based on the output of steel

ingots and castings plus forge and foundry pig iron and the crude steel equivalent of imports less the crude steel equivalent of exports The index is on a daily basis and seasonal variations are eliminated

9 *Consumption of Cotton* Total deliveries of cotton as published by the Liverpool Cotton Association are converted into lb placed on a daily basis and corrected for seasonal fluctuations

10 *Imports of Raw Materials* The value of United Kingdom imports of raw materials less re exports is corrected by the Board

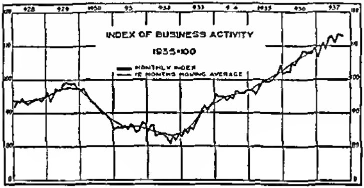


Fig 45

of Trade quarterly index of prices The figures are placed on a daily basis and seasonal variations are eliminated

11 *Exports of British Manufactures* The value figures are corrected for price changes by means of the Board of Trade quarterly index and are placed on a daily basis Seasonal fluctuations are eliminated

12 *Shipping Movements* Based on a daily average of the shipping tonnage engaged in overseas trade entered and cleared at ports in the United Kingdom with seasonal variations removed

13 *Metropolitan Country and Provincial Bank Clearings* Based on daily average clearing figures corrected for price changes by an index constituted as follows Ministry of Labour Index of Wages (2 weights) Ministry of Labour Cost of Living Index (2 weights) *Economist* Index of Wholesale Prices (1 weight) Seasonal variations are removed

...u fon and ear

14. *Town Clearings.* Calculated from the monthly town clearing figures by the same means as in the previous index.

The index is published monthly in the *Economist Trade Supplement*, which contains in addition reports on the current state of trade, and tables of statistical information.

Specimens are shown in Table 108 and Fig. 45.

Other Sources.

Statistical Surveys of business conditions appear in the financial press and the monthly reviews of the leading banks. In addition, there are various economic services conducted upon a commercial basis.

CHAPTER XXX

MISCELLANEOUS APPLICATIONS

Business Statistics

By Business Statistics we understand the application of Statistical Methods to the activities of the individual joint stock or private business enterprise The following headings indicate the scope of this division of the subject—

1 Purchases
2 Production and labour
3 Sales
4 Forecasting
5 Budgeting
6 Consumers purchasing power and market analysis
7 Investment and real estate analysis
8 Finance and credit analysis
9 Control and management statistics
10 External statistics and study of the general economic position

In view of the erratic tendencies of the data Business Statistics afford little scope for the more elaborate methods discussed in Part I The following is a list of methods generally favoured—

1 Analysis of totals into their components

2 Comparisons of current results with budgets or corresponding results for previous periods

3 Comparisons of cumulative results as above cumulating from the beginning of the financial year

4 Reduction of figures to percentages

5 Calculation of averages ratios and moving annual totals (see page 333)

6 Bar charts

7 Graphs

As an example of method the reader should contrast Tables 109 and 110 with the orthodox Balance Sheet and Profit and Loss

TABLE 110

X COMPANY—INCOME STATEMENT

	Last Year			This Year			Increase or Decrease		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Physical Volume of Sales		—	—		—	—		—	—
	£	£ per unit	%	£	£ per unit	%	£	£ per unit	
Gross Sales *Less* Outward Freight Allowances and Returns									
Net Sales *Less* Cost of Goods Sold									
Gross Profit on Sales *Less* Selling Expenses									
Less Administration Expenses (specify)									
Net Profit on Sales (income from operations) Other Income (specify)									
Less Outgo									
Net Income Special Credits to Profit and Loss									
Less Special Debits									
Amount available *Less* Fixed Charges									
Amount Available for Distribution									
Appropriation— Transfers to Reserves Provision for Income Tax Dividends— On Preference Capital On Ordinary Capital Difference in Carry Forward									

TABLE 109

X COMPANY—STATEMENT OF FINANCIAL POSITION

ASSETS

Item	Last Year		This Year		Increase or Decrease
(1)	(2)	(3)	(4)	(5)	(6)
	£	%	£	%	£
Floating Assets—					
Cash					
Sundry debtors					
Stocks					
Marketable securities					
Prepaid expenses					
Other items					
Investments—					
Controlled companies					
Shares and debentures					
Loans					
Other investments					
Fixed Assets—					
Land					
Buildings					
Plant and equipment					
Other items					
Intangible Assets—					
Goodwill					
Patents etc					

LIABILITIES AND NET WORTH

Item	Last Year		This Year		Increase or Decrease
(1)	(2)	(3)	(4)	(5)	(6)
	£	%		%	£
Floating Liabilities—					
Sundry creditors					
Loans					
Specific reserves					
Other items					
Fixed Liabilities—					
Debentures					
Other items					
Net Worth—					
Share capital					
Preference					
Ordinary					
Capital reserves					
Free reserves					
Profit and loss balance					

Account. (This is based upon an American model in which, according to the usual practice the Assets and Liabilities sides of the Balance Sheet are reversed.)

Owing to the variety of problems involved under the heading of Business Statistics, a summary treatment would fail in its purpose, whilst detailed treatment is prevented by lack of space. There are several good all-round introductions to the subject, such as Riggleman and Frisbee's *Business Statistics*; whilst the more ambitious student will find his requirements met in the *Cost and Production Handbook* (edited by L. P. Alford).

The rest of this chapter deals with sundry methods, mainly useful in connection with Business Statistics, which have not already been discussed.

Weighted Totals.

Sometimes it is desired to institute comparisons with regard to the physical volume of business transacted at different periods in varied descriptions of products, irrespective of sterling values. Such comparisons will be particularly valuable during periods of rapidly moving prices when sterling values cease to afford a reliable indication of the amount of business activity. Such a comparison may be effected by the device of weighted totals.

TABLE 111

Z COMPANY—PHYSICAL VOLUME OF TRADING

Product (1)	No. of Units Sold, 1930 (2)	Average Price per Unit (3)	No. of Units Sold 1931 (4)	Average Price per Unit (5)	Average of Cols. (3) and (5) (6)	Weights Proportional to (6) (7)	Col (2) × Col. (7) (8)	Col. (4) × Col. (7) (9)
A	1,000	10·7	900	8·1	9·4	1	1,000	900
B	800	50·9	400	60·7	55·8	5	4,000	2,000
C	50	80·6	60	70·2	75·4	7	350	420
D	600	120·0	500	90·4	105·2	10	6,000	5,000
E	100	40·2	80	35·4	37·8	4	400	320
							11,750	8,640

Upon this basis the total physical volume of sales has been reduced in the ratio 11,750 : 8,640 = 100 : 73·5.

The above operation involves a Quantity Index Number. It

establishes a ratio of quantities weighted by prices instead of a ratio of prices weighted by quantities The formula is—

$$Q_{01} = \frac{\Sigma \frac{1}{2}(p_0 + p_1)Q_1}{\Sigma \frac{1}{2}(p_0 + p_1)Q_0} \qquad (1)$$

Cost and Profit Variation Formulae

It is frequently required to compare actual with budgetary results analysing the variations according to the respective factors involved and the object of these formulae is to provide systematic means of breaking down the total variation into its parts Consider the following data extracted from the books of a firm that has adopted a Standard Costing System

JOB X

Standard labour time	150 hours
Actual	200 hours
Standard hourly rate	1s $5\frac{1}{2}$d
Actual	1s 8d [1]

Total cost variation = 200 × 20 − 150 × 17 5 = 1375d = £5 729 Required to break it down into variations due to (1) rate variation and (2) time variation

According to G Charter Harrison the proper formula is—

Total Cost variation

= actual hours × difference in rates + standard rate × difference in hours

$= 200\ (20 - 17\ 5) + 17\ 5\ (200 - 150)$

$= 500 + 875 = 1375\text{d} \qquad (2)$

This represents a weighted total in which the first weight is based on *actual* and the second on *standard* figures The critical reader will ask how this particular arrangement can be defended and whether it would not be just as reasonable to turn the formulas the other way about making

Total Cost variation

= standard hours × difference in rates + actual rate × difference in hours

$= 150\ (20 - 17\ 5) + 20\ (200 - 150)$

$= 375 + 1000 = 1375\text{d} \qquad (3)$

[1] Adapted from *Cost and Production Handbook* (pp 1197–8)

The explanation is that a sum equivalent to $(200 - 150)(20 - 17{\cdot}5) = 125$d. does not belong entirely to rates nor to hours and that it can only be apportioned according to some convention. If we take the mean of (2) and (3) we have—

$$\tfrac{1}{2}(200 + 150)(20 - 17{\cdot}5) + \tfrac{1}{2}(20 + 17{\cdot}5)(200 - 150)$$
$$= 437{\cdot}5 + 937{\cdot}5 = 1375\text{d.} \qquad (4)$$

Algebraic Analysis

Let $H = AB$ refer to the standard and $H' = A'B'$ to the actual figures. It is required to express the difference $H' - H$ in the form

$$x(A' - A) + y(B' - B)$$

where x and y are to be determined.

Putting $x = B' y = A$ we have

$$A'B' - AB = B'(A' - A) + A(B' - B) \qquad (5)$$

This is Harrison's formula.

Putting $x = B\ y = A'$ we have

$$A'B' - AB = B(A' - A) + A'(B' - B) \qquad (6)$$

This is the alternative formula.

Putting $x = \tfrac{1}{2}(B' + B)$ and $y = \tfrac{1}{2}(A' + A)$ we have

$$A'B' - AB$$
$$= \tfrac{1}{2}(B' + B)(A' - A) + \tfrac{1}{2}(A' + B)(B' - B) \qquad (7)$$

No. (7) is symmetrical, i.e. it is not affected by an interchange between the A's and B's. This advantage is not enjoyed by the other two.

Example. Consider the following data. We budget for the sale of 10,000 tons of Product X at an average price of £10 per ton, utilizing 5,000 composite units of materials and labour at £12 per unit, with overheads totalling £20,000. We sell 12,000 tons at £9·5 per ton, utilizing 5,500 units at £13 per unit, and overheads total £21,000.

Budgeted profit $= 10{,}000 \times 10 - (5{,}000 \times 12 + 20{,}000)$
$= £20{,}000$

Actual profit $= 12{,}000 \times 9{\cdot}5 - (5{,}500 \times 13 + 21{,}000)$
$= £21{,}500$

Difference $= £1{,}500$

Breakdown—

Difference in Profit due to—	£
Increase in quantity sold $= \frac{1}{2}(10 + 9\ 5)(12\,000 - 10\,000)$	19 500
Decrease in selling price $= \frac{1}{2}(10\,000 + 12\,000)(9\ 5 - 10)$	*5 500*
Increase in material and labour utilization $= -\frac{1}{2}(12 + 13)(5\,500 - 5\,000)$	*6 250*
Increase in cost $= -\frac{1}{2}(5\,500 + 5\,000)(13 - 12)$	*5 250*
Increase in overheads $= -(21\,000 \quad 20\,000)$	*1 000*
	£1 500

Analysis—Three Factors

With three factors instead of two the symmetrical formula becomes somewhat complicated viz —

$$
\begin{aligned}
&A\ B\ C - ABC \\
&\quad - \tfrac{1}{6}(2B\ C\ + B\ C + BC\ + 2BC)(A' - A) \\
&\quad + \tfrac{1}{6}(2C\ A\ + C\ A + CA\ + 2CA)(B\ - B) \\
&\quad + \tfrac{1}{6}(2A\ B\ + A\ B + AB\ + 2AB)(C\ - C)
\end{aligned} \tag{8}
$$

Example 1

$$
\begin{aligned}
&A\ = 8\ \ B\ = 4\ \ C\ = 10\ \ A\ B\ C\ - 320 \\
&A\ = 6\ \ B\ = 5\ \ C\ = 7\ \ ABC\ - 210 \\
&\qquad 320 \quad 210 \\
&\quad = \tfrac{1}{6}(80 + 28 + 50 + 70)(8 - 6) \\
&\quad + \tfrac{1}{6}(160 + 60 + 56 + 84)(4 \quad 5) \\
&\quad + \tfrac{1}{6}(64 + 40 + 24 + 60)(10 - 7) \\
&\quad - 76 - 60 + 94 = 110
\end{aligned} \tag{9}
$$

Example 2

Now consider a more complex case Assume

$$P = Qp\left(1 - \frac{s}{100}\right) \quad \left\{Mq + Hw\left(1 + \frac{r}{100}\right) + Ov\right\}$$

The following is a key to the equation—

Symbol	Item	Unit	Budgeted Figure	Actual Figure
Q	Quantity sold	No	10 000	12 000
p	Average selling price	£	20	19
s	Selling expenses	% on sales	15	12
M	Material used	ton	1 000	1 100
q	Price of material	£ per ton	30	32
H	Direct labour	hour	1 000 000	1 150 000
w	Average wage	s d per hour	1s 3d	1s 8d
r	Bonus	% on earnings	5	6
O	Oncosts	as selected	100 000	105 000
v	Oncost rate	s per unit	10	9

We have

$$P = 10{,}000 \times 20 \times \left(1 - \frac{15}{100}\right)$$

$$- \left\{1{,}000 \times 30 + 1{,}000{,}000 \times \frac{15}{240} \times \left(1 + \frac{5}{100}\right) + 100{,}000 \times \frac{10}{20}\right\}$$

$$= 170{,}000 - \{30{,}000 + 65{,}625 + 50{,}000\} = £24{,}375$$

$$P' = 12{,}000 \times 19 \times \left(1 - \frac{12}{100}\right)$$

$$- \left\{1{,}100 \times 32 + 1{,}150{,}000 \times \frac{20}{240} \times \left(1 + \frac{6}{100}\right) + 105{,}000 \times \frac{9}{20}\right\}$$

$$= 200{,}640 - \{35{,}200 + 101{,}583 + 47{,}250\} = £16{,}607$$

$$P' - P = 30{,}640 - \{5{,}200 + 35{,}958 - 2{,}750\}$$

$$= \underline{\underline{£7{,}768}}$$

The analysis may be split into four parts, two involving products of three factors, and the rest products of two factors.

(1) Putting $Q = A$, $p = B$ and $\left(1 - \frac{s}{100}\right) = C$ and using Formula (8), we have

$$A' = 12{,}000,\ B' = 19,\ C' = 0{\cdot}88,\ A'B'C' = 200{,}640$$

$$A = 10{,}000,\ B = 20,\ C = 0{\cdot}85,\ ABC = 170\ 000$$

$$200{,}640 - 170{,}000$$

$$= \tfrac{1}{6}\,(33{\cdot}44 + 16{\cdot}15 + 17{\cdot}60 + 34{\cdot}00)\,(12{,}000 - 10{,}000)$$

$$+ \tfrac{1}{6}\,(21{,}120 + 8{,}800 + 10{,}200 + 17{,}000)\,(19 - 20)$$

$$+ \tfrac{1}{6}\,(456{,}000 + 240{,}000 + 190{,}000 + 400{,}000)\,(0{\cdot}88 - 0{\cdot}85)$$

$$= 33{,}730 - 9{,}520 + 6{,}430 = £30{,}640 \quad . \quad . \quad . \quad . \quad (10)$$

(2) Putting $M = A$, $q = B$ and using formula (7) we have—

$$A' = 1100,\ B' = 32,\ A'B' = 35{,}200$$

$$A = 1000,\ B = 30,\ AB = 30{,}000$$

$$35{,}200 - 30{,}000$$

$$= \tfrac{1}{2}\,(32 + 30)\,(1100 - 1000) + \tfrac{1}{2}\,(1100 + 1000)\,(32 - 30)$$

$$= 3100 + 2100 = £5200 \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

(3) Putting $H = A$ $w = B\left(1 + \frac{r}{100}\right) = C$ and using formula (8) we have—

$$A = 1\,150\,000 \quad B = \tfrac{1}{12} \quad C = 1.06 \quad ABC = 101\,583$$
$$A = 1\,000\,000 \quad B = \tfrac{1}{16} \quad C = 1.05 \quad ABC = 65\,625$$

$$101\,583 - 65\,625$$
$$= \tfrac{1}{6}(0.17667 + 0.08750 + 0.06625 + 0.13125)(1\,150\,000 - 1\,000\,000)$$
$$+ \tfrac{1}{6}(2\,438\,000 + 1\,060\,000 + 1\,207\,500 + 2\,100\,000)(\tfrac{1}{12} - \tfrac{1}{16})$$
$$+ \tfrac{1}{6}(191\,667 + 71\,875 + 83\,333 + 125\,000)(1.06 - 1.05)$$
$$= 11\,542 + 23\,630 + 786 = £35\,958 \qquad (12)$$

(4) Putting $O = A$, $v = B$ and using formula (7) we have

$$A = 105\,000 \quad B = \tfrac{9}{20} \quad AB = 47\,250$$
$$A = 100\,000 \quad B = \tfrac{1}{2} \quad AB = 50\,000$$

$$47\,250 - 50\,000$$
$$= \tfrac{1}{2}(\tfrac{9}{20} + \tfrac{1}{2})(105\,000 - 100\,000) + \tfrac{1}{2}(105\,000 + 100\,000)(\tfrac{9}{20} - \tfrac{1}{2})$$
$$= 2375 - 5125 = £2\,750 \qquad (13)$$

Collecting the results we have—

Cause of Variation	Equation No	Amount of Variation
		£
Increase in quantity sold	(10)	33 730
Decrease in price		*9 520*
Decrease in selling expenses		6 430
		30 640
Increase in material usage	(11)	*3 100*
Increase in price		*2 100*
		5 200
Increase in labour usage	(12)	*11 542*
Increase in wage rates		*23 630*
Increase in bonus		*786*
		35 958
Increase in oncost units	(13)	*2 375*
Decrease in rates		*5 125*
		2 750
TOTALS		*7 768*

The calculations are laborious but straightforward. It would not of course be worth while going to all this trouble without a special object in view.

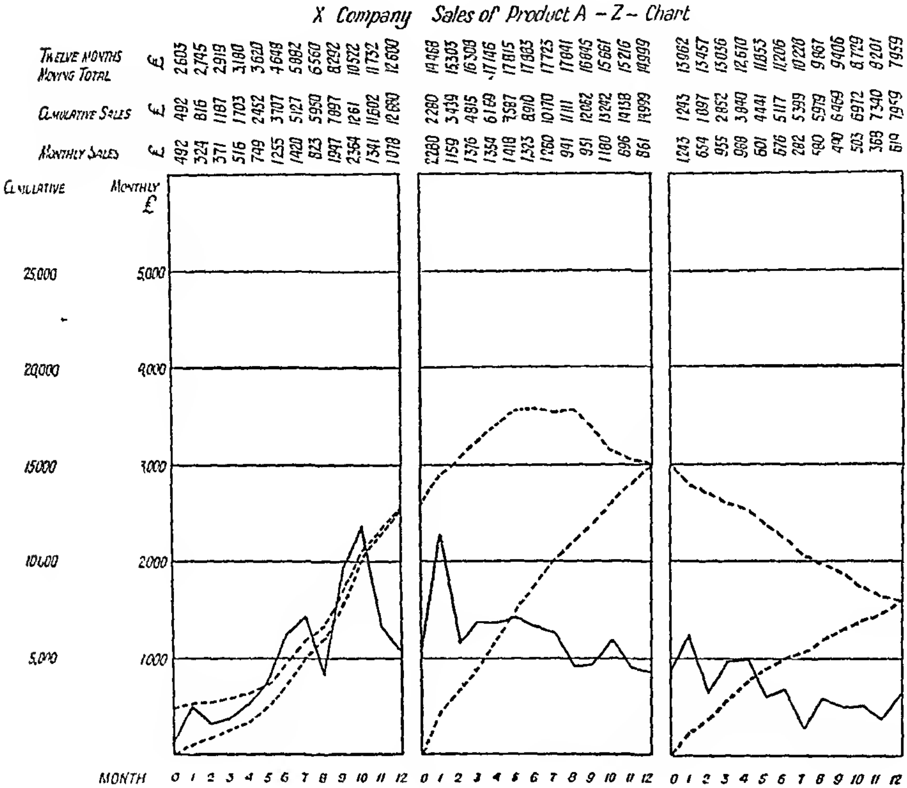


FIG. 46

Z-charts.

The Z-chart gains its name from the fact that its three curves roughly form the letter "Z." It is an historical chart whose elements comprise—

1. The curve of the original data.
2. A cumulative curve of the above.
3. A moving total curve.

The cumulative curve shows the progress to date from the beginning of the statistical period. It smoothes out irregularities and is especially valuable in those classes of business in which an excess

(deficiency) in one month is likely to be compensated by a deficiency (excess) later on. By studying the direction of the curve relatively to that taken the previous year it is possible to make a fair forecast of results for the year

The moving annual total (M.A.T.) has the following interesting property—

> M.A.T. for current month = M.A.T. for last month + difference between figures for current month and those for corresponding month of previous year

The difference in question is affected by accidental variations and one should not take too much notice of individual movements. When however this difference taken over several months shows persistent decrease (indicated by a slackening in the *rate* of advance by the M.A.T. curve) it is a sign of weakness although the curve may be still rising

The great advantage of the M.A.T. curve is that it eliminates seasonal fluctuations with a minimum of labour

It is not necessary to plot all three curves. For many purposes the M.A.T. curve above will suffice

The Z-chart can be adapted to any period desired but the most common variants are—

(1) Monthly figures	12 to the year
(2) Periodic	13
(3) Weekly	52
(4) Daily	28 to a period or 30 or 31 to a month

In the first three cases the moving total is an annual one and in the last a monthly one

Fig 46 shows Z-charts extending over a period of three years placed side by side. There are two scales one for the monthly and the other for the cumulative figures the latter being five times the former. The firm line indicates the monthly figures and the dotted lines the other two [1]

The cumulative figures show the totals for the year to date beginning 1st January and the moving total figures the cumulation for the past twelve months. The cumulative and moving total figures meet at 31st December of each year (as they should)

[1] In practice the monthly figures and their relative scales would be plotted in black and the cumulative figures and their relative scales in red

With weekly figures the cumulative scale should be twenty times the weekly scale, and with daily figures ten times the daily scale.

Progress Charts.

The object of a Progress Chart is to keep a continuous and up-to-date record of actual progress against some pre-determined standard. The following illustration relates to a Chart on the Gantt system.

The Gantt System.

This system expresses actual performance as a percentage of a monthly quota. The chart is not scaled according to physical quantities or values but according to Time.

Suppose we are carrying out a sales or constructional programme, and that the quotas for the twelve months of the year are as shown in Table 112.

The first step is to prepare a blank chart as indicated in Fig. 47. The names of the items are inserted on the left-hand side and the names of the months along the top. The quota figure for each month is then entered in the north-west corner of each compartment, and the total quota to date in the north-east corner.

TABLE 112

ILLUSTRATING THE GANTT CHART

Month (1)	Quota Figures (2)	Cumulative Figures (3)
January	360	360
February	420	780
March	420	1,200
April	640	1,840
May	820	2,660
June	580	3,240
July	340	3,580
August	740	4,320
September	600	4,920
October	560	5,480
November	500	5,980
December	420	6,400

GANTT PROGRESS CHART

These Figures represent the same Chart at successive stages

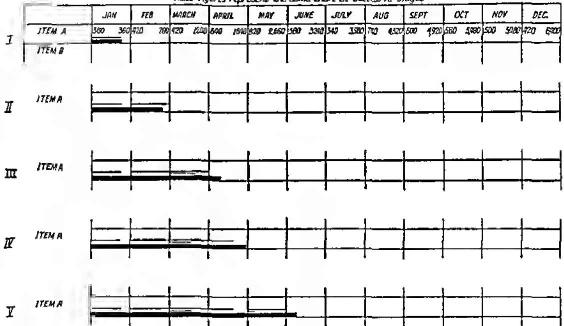


FIG 47

We shall now trace out the progress of the chart month by month under item A.

MONTHLY PERFORMANCE

Month (1)	Quota (2)	Performance (3)	As Percentage of Quota (4)	How Plotted (5)
January .	360	270	75	Thin line covering 75 per cent of column.
February .	420	420	100	Thin line covering 100 per cent of column.
March . .	420	710	169	Thin lines covering 100 per cent plus 69 per cent of columns.
April . .	640	400	62	Thin line covering 62 per cent of column.
May . .	820	1,000	122	Thin lines covering 100 per cent plus 22 per cent of columns.

CUMULATIVE PERFORMANCE

Month (1)	Quota to Date (2)	Performance to Date (3)	Analysis of (3) (4)	How Plotted (5)
January .	360	270	75 per cent of quota.	Thick line covering 75 per cent of column.
February .	760	690	360 to provide January quota plus 330 on account of February quota = 79 per cent.	Thick line covering January column plus 79 per cent of February column.
March . .	1,200	1,400	1,200 to provide up to end March plus 200 on account of April quota = 31 per cent.	Thick line covering January-March column plus 31 per cent of April column.
April . .	1,840	1,800	1,200 to provide up to end March plus 600 on account of April quota = 94 per cent.	Thick line covering January-March columns plus 94 per cent of April column.
May . .	2,660	2,800	2,660 to provide up to end May plus 140 on account of June quota = 24 per cent.	Thick line covering January-May columns plus 24 per cent of June column.

APPENDIX II

REFERENCES TO PART II

OFFICIAL STATISTICS

AGRICULTURE AND FISHERIES, MINISTRY OF—
Journal of the Ministry of Agriculture (monthly)
Agricultural Statistics (annually)
Agricultural Returns of England and Wales (annually)
Agricultural Output of England and Wales 1930–31

CUSTOMS AND EXCISE BOARD OF—
Report of Commissioners of H M Customs and Excise (annually)

EDUCATION BOARD OF—
Various

ELECTRICITY COMMISSION—
Various

GENERAL REGISTER OFFICE—
Returns of Births and Deaths (weekly and quarterly)
Registrar General s Statistical Review of England and Wales (2 vols annually)
Registrar General s Decennial Supplement 1931
Census of England and Wales 1931—
Preliminary Report
County Volumes
Occupation Tables
Industry Tables
Housing
General Tables Population Institutions Ages and Marital Conditions Birthplace and Nationality
General Report with Appendices

HEALTH MINISTRY OF—
Number of Persons in Receipt of Poor Law Relief (quarterly)
Report of Ministry (annually)
Local Rates—Poundage and Valuations (annually)
Local Taxation Returns (annually)
National Health Insurance Fund Accounts (annually)
Housing House Production Slum Clearance etc (half yearly)
Report on Overcrowding Survey England and Wales

HOME OFFICE—
Criminal Statistics (annually)
Road Accidents Involving Personal Injury (annually)

IMPERIAL INSTITUTE MINERAL RESOURCES DEPARTMENT—
The Mineral Industry of the British Empire and Foreign Countries (annually)

INLAND REVENUE BOARD OF—
Report of Commissioners of Inland Revenue (annually)

LABOUR MINISTRY OF—
Ministry of Labour Gazette (monthly)
Abstract of Labour Statistics (every three years)

LORD CHANCELLOR'S DEPARTMENT—

Civil Judicial Statistics (annually).

MINES DEPARTMENT—

Statistical Summary of Output and Costs of Production, Proceeds and Profits of the Coal Mining Industry (quarterly).

OVERSEAS TRADE (DEVELOPMENT AND INTELLIGENCE), DEPARTMENT OF—

Reports by H.M. Trade Commissioners.

TRADE, BOARD OF—

Board of Trade Journal (weekly).
Accounts Relating to the Trade and Navigation of the United Kingdom (monthly).
Foreign Trade and Commerce (quarterly).
Annual Statement of Trade of United Kingdom, 4 vols. (annually).
Navigation and Shipping, Annual Statement.
Statistical Abstract for the United Kingdom (annually).
Fourth Census of Production, 1930—
Final Reports.
Fifth Census of Production, 1935, Preliminary Reports.

TRANSPORT, MINISTRY OF—

Railway Statistics (monthly).
Railway Companies (Staff) (annually).
Road Traffic Census.

TREASURY—

Staffs employed in Government Department (annually).
Consolidated Fund (Abstract) (annually).
Account of Public Income and Expenditure of the United Kingdom (annually).
Finance Accounts of the United Kingdom (annually).
Financial Statement (Budget) (annually).
National Debt Returns (annually).
Public Social Services Return (annually).
Civil Estimates and Estimates for Revenue Departments (annually).

DEPARTMENTAL AND OTHER COMMITTEES, ROYAL COMMISSIONS, ETC.—

Committee on Finance and Industry—Report (Cmd. 3897), 1931.
Final Report of the Committee on Industry and Trade (Cmd. 3282), 1929.
Committee on Industry and Trade, Survey of Industrial Relations, 1926.
Committee on Industry and Trade: Factors in Industrial and Commercial Efficiency (1927).
Committee on Industry and Trade: Further Factors in Industrial and Commercial Efficiency (1928).
Report of the Committee on National Debt and Taxation (Cmd. 2800), 1927.
Committee on National Expenditure Report (Cmd. 3920), 1931.
League of Nations: Monthly Bulletin of Statistics, passim. (Messrs. Constable are the authorized agents for Great Britain.)

PRIVATE PUBLICATIONS—

Allen, R. G. D., and Bowley, A. L.: *Family Expenditure* (King, London, 1935).
Bliss, J. H.: *Financial and Operating Ratios in Management* (Ronald, New York, 1923).
Bowley, A. L.: *An Elementary Manual of Statistics* (Macdonald & Evans, London, 1928).
Bowley, A. L.: *Elements of Statistics* (King, London, 1937).

PRIVATE PUBLICATIONS—(*contd*)

Bowley A L *Official Statistics and What They Contain and How to Use Them* (H Milford 1928)

Bowley A L *Nature and Purpose of the Measurement of Social Phenomena* (King London 1923)

Bowley A L. and Hogg M H *Has Poverty Diminished A Sequel to Livelihood and Poverty* (King London 1925)

Carr Saunders A M. and Jones C D *Survey of the Social Structure of England and Wales* (Oxford 1937)

Chisholm C (Editor) *Marketing Survey of the United Kingdom* (Business Publications London 1936)

Clark Colin *National Income and Outlay* (Macmillan Cambridge 1937)

Daniels G W and Campion H *The Distribution of National Capital* (Manchester University Press, 1936)

Dent A G H *Management Planning and Control* (Gee London 1935)

Fisher Irving *The Making of Index Numbers* (H Mifflin Boston 1922)

Haberler G von *Prosperity and Depression* (League of Nations Geneva 1937)

Harrison G and Mitchell F C *The Home Market* (Unwin London 1936)

Hill A Bradford *Principles of Medical Statistics* (*The Lancet* London 1937)

Jones D Caradog (Editor) *The Social Survey of the Merseyside* (University Press Liverpool 1934)

Karsten K G *Charts and Graphs* (Prentice Hall New York)

Kuznets S S *Seasonal Variations in Industry and Trade* (National Bureau of Economic Research New York 1933)

Kuznets S S *Secular Movements in Production and Prices Their Nature and Bearing upon Cyclical Fluctuations* (H Mifflin New York 1930)

Layton Sir W and Crowther G *An Introduction to the Study of Prices* (Macmillan London 1935)

Mitchell W C *Business Cycles the Problem and its Setting* (National Bureau of Economic Research New York 1927)

New Survey of London Life and Labour (King London 1930)

Newsholme A *Elements of Vital Statistics* (Allen & Unwin London 1924)

Orr J B *Food Health and Income* (Macmillan London 1936)

Pearson E S *The Application of Statistical Methods to Industrial Standardisation and Quality Control* (British Standards Institution London 1935)

Persons W M *Forecasting Business Cycles* (Wiley New York 1931)

Persons W M and others *Problems of Business Forecasting* (Mifflin Boston 1924)

Riggleman J R and Frisbee I N *Business Statistics* (McGraw Hill 1932)

Rose T G *Business Charts* (Pitman London 1937)

Rose T G *Higher Control* (Pitman London 1934)

Rowntree B Seebohm *The Human Needs of Labour* (Longmans London 1937)

Saunders C *Seasonal Variations in Employment* (Longmans, London 1936)

Secrist H *An Introduction to Statistical Methods* (Macmillan New York 1925)

Shirras G Findlay *The Science of Public Finance* 2 vols (Macmillan London 1936)

Stamp Sir J C *British Incomes and Property—The Application of Official Statistics to Economic Problems* (King London 1916)

Stamp Sir J C *Current Problems in Finance and Government* (King London 1924)

Stamp Sir J C *The Fundamental Principles of Taxation in the Light of Modern Developments* (Macmillan London 1936)

Stamp, Sir J. C.: *Some Economic Factors in Modern Life* (King, London, 1929).

Stamp, Sir J. C.: *Wealth and Taxable Capacity* (King, London, 1922).

Statesman's Year Book, edited by M. Epstein (Macmillan, London) (annually).

Schultz, H.: *Statistical Laws of Supply and Demand, with special application to Sugar* (University of Chicago Press, 1928).

Verwey, G.: *The Economist's Handbook, a Manual of Statistical Sources* (Amsterdam, 1934).

Wallace, W.: *Business Forecasting and its Practical Application* (Pitman, London, 1927).

Wedgwood, Josiah: *The Economics of Inheritance* (Routledge, London, 1929).

Willsmore, A. W.: *Business Budgets and Budgeting Control* (Pitman, London, 1936).

PERIODICALS—

The Accountant. The recognized weekly organ of Chartered Accountants and Accountancy throughout the world (Gee, London).

The Economic Journal. The Journal of the Royal Economic Society (Macmillan).

The Economist, 8 Bouverie Street, E.C.4.

Journal of the Royal Statistical Society. The Royal Statistical Society, 4 Portugal Street, W.C.2.

APPENDIX III

CALCULATING MACHINES

By L J COMRIE MA PhD

Late Superintendent H M Nautical Almanac Office

WHEN the scale of arithmetical operations becomes large it is the modern practice to employ mechanical aids to computation The principal advantages of such aids may be enumerated thus—

(1) They enable investigations that would otherwise be beyond the capacity of human beings to be undertaken

(2) They enable results to be made available in time for action based on them to be effective

(3) They enable a great deal of routine work to be done by inexpensive labour without excessive drudgery thus releasing the time of trained investigators for the more skilled work of collecting data deciding to wha processes they shall be subjected and interpreting the results of those processes

(4) They assist in reducing the cost of computation to the point where it is only a fraction of the value of the derived information

(5) They eliminate a great deal of the writing that is usually necessary in computation and hence remove one of the most frequent sources of error An excellent example of this occurs in the formation of the sum of a number of products the individual products are added as they are formed

We shall proceed to examine briefly the main types of machine available with some indication of the arithmetical field in which each would be used

Adding and Listing Machines

Although multiplying machines (see below) may be used for addition and subtraction and are so used when only one general purpose machine can be afforded it is preferable if there is much work of this nature to be done to use a machine specially constructed for fast addition and subtraction with recording of items entered and their totals

Full Keyboard Machines.

In this type, exemplified by the Burroughs, Continental (Fig 48), National (Fig. 53), and Victor (Fig. 49), there is a bank of keys numbered 1 to 9 (or 11 in a pence column) for each denomination in

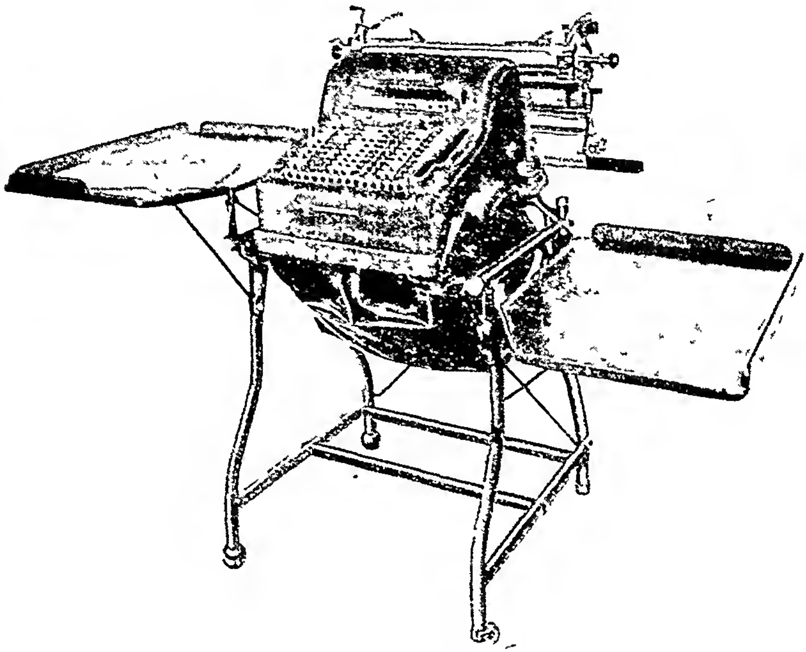


FIG 48. CONTINENTAL ADDING AND ACCOUNTING MACHINE

the amounts to be added. The appropriate keys are depressed, and a motor-bar pressed in electric models or a handle pulled in hand models. This operation prints the number and causes it to be added or subtracted, according to the setting of the controls. When all the items are entered, a *sub-total* key enables the sum to be printed and retained in the machine, or a *total* key prints the sum and clears or zeroises the adding mechanism.

Ten-key Machines.

These machines, typified by the Sundstrand (Fig. 50), Remington and Burroughs Typewriter Accounting machine (Fig. 52), differ only in the method of setting numbers There are ten keys only in a decimal machine, or twelve in a sterling machine. The keys

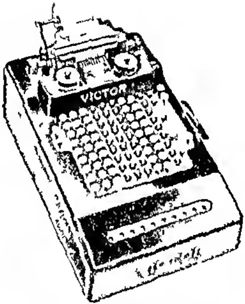


Fig 49 Victor Adding and Listing Machine

Fig 50 Sundstrand 10 Key Adding and Listing Machine

corresponding to the amount to be entered are struck in succession, as on a typewriter. At every stage of this setting the figure last struck is in the units position. When the setting is complete, the motor-bar is depressed to cause the printing, with addition or subtraction.

Each type has its merits; neither is outstandingly superior to the other. For the full keyboard machine it is claimed that the number can be checked before printing, that several keys can be set at once, that ciphers do not have to be set, and that a large machine can be split into two or more fields easily, especially with the aid of a colour scheme in the keyboard and an optional cipher split, which (when in use) suppresses ciphers between the point where it is effective and the next significant figure. With ten-key machines no selection of columns is required, and a touch system can be developed. It may perhaps be said that the full-keyboard type is, on the whole, best suited to general work and casual operators, but that the ten-key type, if applied in work where all the conditions are favourable, may be found more suitable.

Key-driven Machines.

In these machines, represented by the Burroughs Calculator, the Felt and Tarrant Comptometer, and the Plus, addition is performed by the simple depression of keys. There is thus no recording of the amounts entered, so that repetition is usually necessary as a check. Subtraction is performed (except on a recent two-register Burroughs model) by the artificial process of adding complements. In the hands of highly-trained and skilled operators these machines are faster than any other manually-operated type. But the intensive training and constant practice necessary to attain and maintain proficiency usually result in confining the successful use of these machines to places where the bulk of the work is sufficient to enable a large full-time female staff to be employed.

Typewriter Accounting Machines.

These machines, which include the Elliott Fisher, Mercedes (Fig. 51), Remington, Smith Premier, and Underwood, represent a development of the typewriter, in which an assembly of totalisers is carried on a rack moving with the carriage. The totalisers are actuated by the keys of the typewriter. Thus all the items entered

in any column go into a particular totaliser which at the close of the operations will give the total of the items in that column The last totaliser usually receives positively or negatively all the items in each line and so gives balances line by line, in other words it totals or foots across the line and thus receives the

Fig 51 Mercedes Typewriter Accounting Machine

name *crossfooter* Balances or totals cannot be cleared and printed by a single movement but must be read visually and typed out

*These machines permit the analysis of many accounting opera*tions such as public utility billing payrolls and ledgers Their chief advantage is the ease with which descriptive or literal matter can be included in the record On the other hand they give the impression of being a makeshift i e a typewriter with adding mechanism superimposed as an afterthought They are also subject to the limitation that transfer from one totaliser to another is not possible

Multi-register Machines

The objections of the last paragraph do not apply to the Burroughs Typewriter Book keeping machine (Fig 52) which is a multi register ten key adding and subtracting machine with provision for registers up to 20 Transfer from register to register may be effected via the crossfooter

Other forms of the Burroughs machine, with as many as ten registers and a crossfooter, have a full keyboard, but no typewriter, although three or four columns of the type are usually reserved for descriptive matter. Two forms of the Continental machine (Fig. 48) have similar properties. The Sundstrand (Fig. 50), which has already been mentioned, is available as a four-register machine.

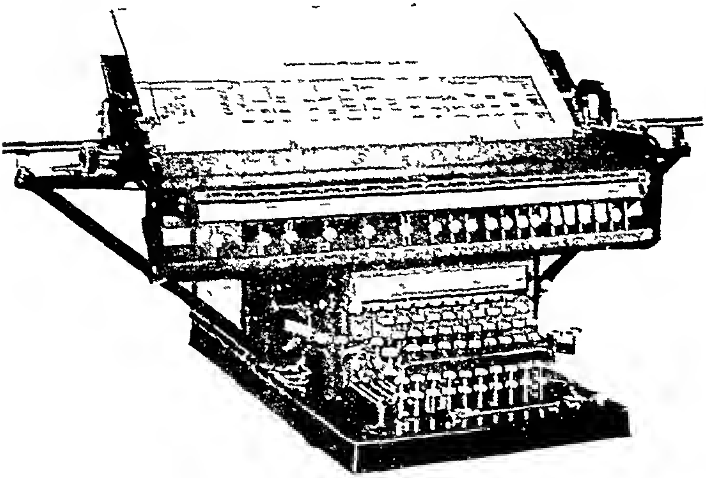


FIG 52 BURROUGHS MULTIPLE-TOTAL TYPEWRITER BOOK-KEEPING MACHINE

The National Cash Register Analysis machine has thirty key-selected registers.

The National Accounting Machine.

This machine (Fig. 53) is really a member of the last-mentioned group, but has properties that lend themselves so well to scientific work that it is described separately. It is a six-register machine, with a typewriter if desired, in which two of the registers will subtract as well as add. Its unique property, on which its scientific utility depends, is that a number set on its twelve-column keyboard may be printed and entered in one operation into any combination of the registers, including subtraction in the two that will subtract; further, a number in any register may be printed and in the same operation transferred (with or without clearing) to any combination of the remaining registers.

As a full description of this machine is available elsewhere it suffices to mention some of the problems to which it has been applied

(1) Mechanical integration with printing from finite differences

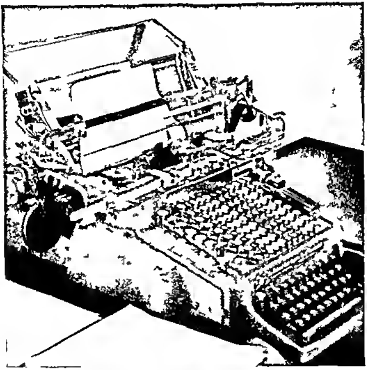


FIG 53 NATIONAL ACCOUNTING MACHINE

up to the sixth This is applicable in the evaluation of polynomials or the solution of differential equations

(2) Differencing of values at regular intervals of the independent variable Differences up to the fifth are obtained and printed by the setting of the function and manipulation of controls higher differences can be obtained by repeating the process with the fifth difference already obtained Differencing is the most powerful method of detecting accidental errors in a series of function values

it is also required as a preliminary stage in interpolation or in subtabulation. It is very laborious when done by hand, and ordinary calculating machines afford only limited help, because of the writing and setting necessary.

(3) Formation of moments by summation, when the intervals are equidistant. First and second moments are required in getting means and standard deviations; they can be obtained by setting the data and operating the controls. Fourth, fifth, and sixth moments are used in various curve-fitting processes.

(4) The formation of the sums of numbers in groups, as required in various smoothing processes.

(5) Subtabulation, or the systematic breaking down of tables to smaller intervals. In doing this it is customary to allow the machine to round off the values produced at the desired point, and use the printed figures as printers' copy.

This machine has lightened the burden of computation in several of the heaviest computing programmes that have been undertaken in the last few years. It seems destined to have a far-reaching influence on table-making in particular.

Calculating Machines

The term " calculating machine " is properly applied to the group of machines that cater primarily for multiplication and division. Most of these (the three exceptions will be mentioned later) are, in truth, only adding and subtracting machines, because multiplication can be performed by continued addition and division by continued subtraction. As facilities for seeing these machines demonstrated are so readily available, they need not here be described in detail, but a rapid survey of the main types and their outstanding features will be made. A good description of the operating principles will be found in *The Office Machine Manual*, Vol. 2, No. 1. (September, 1937.)

Hand-operated Machines.

The most successful surviving hand machines are those of the barrel type, so named from their shape. The best known of these, and the most popular in scientific circles, is the Brunsviga (Fig. 54). The model illustrated has a capacity of twelve setting levers (for the multiplicand), eleven digits in the revolution or multiplier

register and twenty digits in the product register The multiplier register has two sets of figures one in white for forward revolutions and the other in red for backward revolutions An ingenious sliding window that allows one set or the other to be seen adjusts itself automatically according to the nature of the operation This register has tens transmission—a feature that is lacking on older machines and even on some modern ones It is considered essential

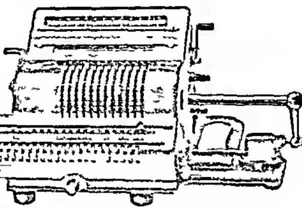


FIG 54 BRUNSVIGA CALCULATING MACHINE MODEL 20

for the proper checking of multiplication in which short-cutting has been used e g the entry of 19 as two forward turns in the tens position and one backward turn in the units position Since short cutting reduces the number of revolutions by more than 40 per cent it cannot be neglected in fact it is because the human operator can short-cut whereas automatic electric machines with one exception (the Hamann Selecta) do not that the margin between hand and electric machines is so small

The most valuable feature which is not shared by any other machine in the same convenient form is the ability to transfer a number from the product register to the setting levers This lends itself immediately to the continued multiplication of three or more factors Another almost equally useful application of this feature is to the conversion of complements in the product register to their direct form this is done by transferring the product to the setting levers and then subtracting it from the zeros in the product register

Another feature is a means of splitting the clearing of the product register, so that the right-hand ten digits are cleared by the clearing lever, while the left-hand ten digits remain. With the aid of this, and the transfer feature just described, it is possible to accumulate the sum of a series of products (whether of two or more factors each), whilst seeing each individual product. This is done by developing products in the right-hand side of the product register,

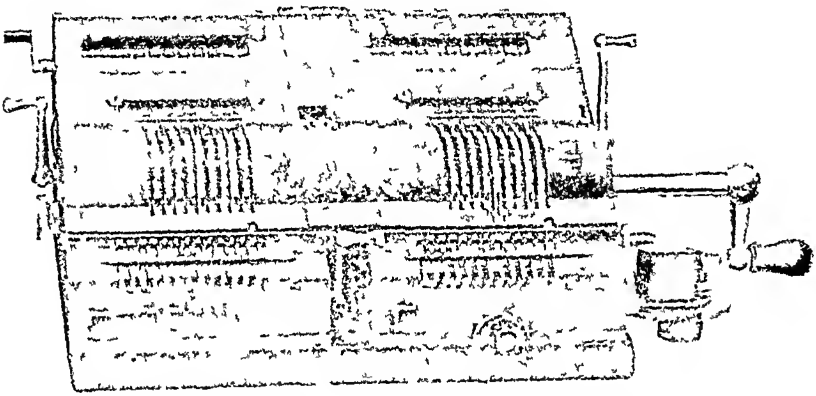


FIG 55 BRUNSVIGA CALCULATING MACHINE, MODEL TWIN 13Z

transferring them to the setting levers, moving the carriage to the extreme right, and then, by turning the handle, storing the product in the non-clearing left-hand side of the product register. This same combination of features also lends itself to mechanical integration from second differences, and to the formation of first and second moments of data at equidistant intervals of the argument.

The large capacity serves not only for handling large numbers but also for handling two small numbers simultaneously. Thus if we want (as in calculating correlations) Σa^2, Σab and Σb^2, we may set $a \times 10^7 + b$, and multiply it by itself, thus giving $a^2 \times 10^{14}$, $2ab \times 10^7$, and b^2. If a and b are not too large, these three products will be separated in the product register.

A variant of this machine, in which two machines, each of capacity $10 \times 8 \times 13$, are formed into a twin machine with a common crank, is shown in Fig. 55. The obvious application is to the multiplication of two numbers by a common multiplier, but many "stunts" are possible, such as the performance of the operation symbolised by

$ab - c$ in a single operation Multipliers may be accumulated in the second multiplier register

An older type of hand machine based on the machine of Thomas de Colmar who in turn used the stepped wheel first designed by Leibnitz is still encountered occasionally The term arithmometer is often applied to this group These machines were characterised by the facts that the handle could be turned in one direction only and that a reversing lever had to be moved to cause a subtraction They are now considered obsolete

Electric Machines

The term electric machine although in common use is misleading as it denotes not a machine that calculates electrically but one that is electrically driven The advantage of the mere replacing of human energy by electrical energy is slight the best electric machines are valued for their auxiliary features—especially automatic multiplication and division Most electric machines have keyboard setting as contrasted with the lever setting of barrel type machines or the slide set up of arithmometers This is an undoubted advantage where a considerable amount of setting is involved as for instance if the machine is also used for adding and subtracting

One of the most highly developed and spectacular electric machines is the Mercedes of which the Model 38 MS is illustrated in Fig 56 In automatic multiplication its bank of sixteen keys is considered to be divided into two halves The multiplier is set on the left hand half and the multiplicand on the right on depressing a multiplier key the multiplication is performed and the multiplier appears in a multiplier register and the product in a product register Once the factors are set multiplication of an eight figure number by another eight figure number takes about eight seconds Automatic division is performed in a similar manner The dividend is set on the left of the keyboard and the divisor on the right the division key is then pressed and the quotient appears in the multiplier register

As the multiplier register is not required for the checking of individual multipliers it may be used for their accumulation controls enable products to be put positively or negatively in the product register at will and multipliers to be stored positively or negatively in the multiplier register Thus with a series of small

numbers a and b, by using $10^7 + a$ as the multiplier and b as multiplicand, the product register will show, at the end of the run, $\Sigma b.10^7 + \Sigma ab$ and the multiplier register will show the last figure of the number of products, as well as Σa. This can be used to evaluate Σa and Σb as well as Σab, or, if the former are known, to check the setting of a and b. Similarly Σa and Σa^2 can be found in one run, with a check on the setting of each value of a. The

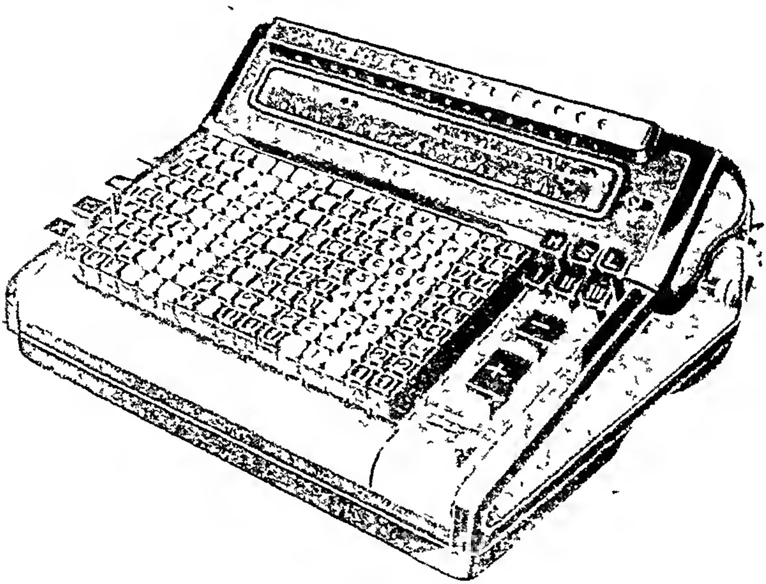


FIG 56 MERCEDES ELECTRIC CALCULATING MACHINE, MODEL 38 MS

value of this feature in correlation work, or in the analysis of variance, is self-evident.

Another feature permits the continued multiplication of three or more factors. Suppose two factors have already been multiplied. On pressing the M key, the left-hand eight digits in the product register are transferred to a temporary holding mechanism. A third factor is then set on the right of the keyboard, and, on pressing the multiplication key, is multiplied by the product that was being held. If successive powers are required, the right of the keyboard is left undisturbed; each pressure on the M and multiplication keys then gives a higher power.

An invisible storage register will accept the contents of the product register restoring them to that register with addition to anything standing there when desired This enables the sum of a series of products to be formed while still permitting examination of the individual products In conjunction with the M key it enables products of any number of factors to be formed seen and accumulated The value of a complement in the product register may be read in direct form by sliding a window which conceals the complement and reveals the direct figures

Muld n To lo

FIG 57 MADAS ELECTRIC CALCULATING MACHINE

The Hamann Selectamachine has two separate keyboards on which multiplier and multiplicand may be set It is unique in that it multiplies with automatic short cutting In the Rheinmetall and Frieden machines the multiplier is set by an auxiliary ten key setting mechanism as soon as the last figure has been set the multiplication key is depressed Provision is made for optional automatic clearing of the multiplier and product registers after this depression and before the new product is begun

In the Madas (Fig 57) and Archimedes machines a single keyboard serves for both multiplier and multiplicand The former is set first transferred by the pressing of a key to a holding mechanism then the multiplicand is set in the usual way and the multiplication key depressed With the Madas machine the use of a repeat key *enables the square of a number set on the keyboard to be formed* by two depressions of the motor bar The sum of a series of numbers

so treated would be accumulated in the multiplier register; this is ideal in the analysis of variance. The contents of the product register may be used as a later multiplier; whereas with the Mercedes this possibility is limited to the contents of the left-hand portion of the product register, this restriction does not exist in the Madas.

The Marchant (Fig. 58) machine receives its multiplier from an auxiliary column of ten keys on the right of the main keyboard.

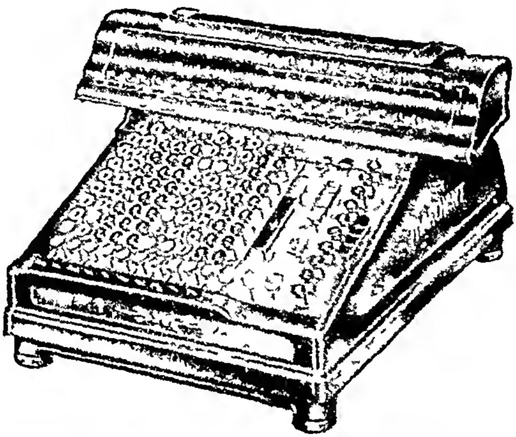


FIG 58 MARCHANT ELECTRIC CALCULATING MACHINE

The depression of any key causes the machine to make the desired number of revolutions, and step the carriage to the next position. While these revolutions are being made, the next key may be depressed in advance; this, combined with the high speed of revolution—900 a minute—serves to make the Marchant fast and easily operated for straightforward multiplications. This machine is much less noisy than most; it is also one of the few machines in which the product register has tens transmission throughout twenty figures.

Another well-known machine is the Monroe, whose special features consist of a second multiplier register (unfortunately without tens transmission) and a second product register, which may be used for accumulating products while the individual products are shown in the main product register.

The Facit machine (Fig. 59), which is really of the barrel type, is unique in that its setting is done by means of a ten-key keyboard.

This facilitates its use for addition and subtraction but the fact that no change of setting can be made without destroying the entire setting is an unfortunate sacrifice of flexibility

It will be observed that the leading advantages of electric machines are their keyboard setting their facilities for automatic multiplication and division and special features such as those for continued multiplication for accumulation of multipliers and for

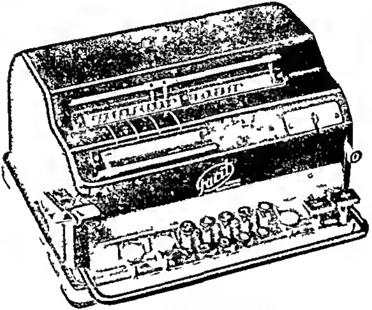


FIG 59 FACIT ELECTRIC CALCULATING MACHINE

storage Nevertheless the day of the hand machine is not past It is naturally less expensive and more easily transported Its advantage derived from short cutting has already been mentioned Also once the multiplicand has been set multiplication may begin whereas in automatic machines (except the Marchant) there is a further delay while the multiplier is being set In a well organised computing establishment—especially for scientific work—it will be found that the use of various types of machine affords greater computing power than standardisation on any one machine

Direct-multiplication Machines.

In the machines already described multiplication is performed by repeated addition. In a direct-multiplication machine a mechanical multiplication table up to 9×9 is incorporated. Hence in multiplication the required digits are selected, and the time required to multiply by any digit is independent of the value of that digit.

The best-known representative of this class is the Millionaire. An arm at the left of the machine sweeps over an arc, and selects the multiplier digits, after which the handle is turned or an electric button pressed. During the revolution the tens of the partial products are first added, then the carriage is stepped one position to the left, and the units of the partial products added. In division the correct digit of the quotient must be estimated before the handle is turned; this offsets the advantage given by the smaller number of handle turns, especially if an error is made in the estimate in borderline cases. This machine, very popular twenty-five to thirty years ago, has now given way to the more modern machines described above.

Certain models of the Burroughs Typewriter Book-keeping machine (Fig. 52), formerly known as the Moon-Hopkins, contain a direct multiplying mechanism. There are two sets of ten entering keys—one for the multiplicand and one for the multiplier. The multiplicands and the leading figures of the products (but not the multipliers) may be printed, and the products may be stored in a series of registers—up to 20 if desired. Although this commercial machine lends itself to certain forms of scientific work, no actual scientific applications are known.

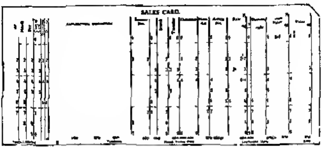
The third member of this group, namely the Hollerith multiplying punch, is described below.

Punched-card Sorting and Tabulating Machines

When the number of classes into which items have to be analysed becomes greater than the number of registers in the multi-register machines already mentioned, or when the same items have to be analysed into various groups, an entirely different principle is resorted to, namely the use of punched cards that can be sorted mechanically and used to actuate adding mechanism.

Information is conveyed to the machines by cards (Fig. 60) containing a series of columns, in each of which any one of the

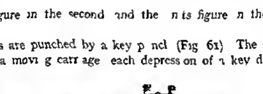
d gits o 1 9 n ay be puncl ed A group of se eral columns s usually described as a field tl us three figure 1 umbers would requ re a three column field the l undreds figure be ng m the first column



F G 60 PUNCHED CARD AS USED N HO LER TH MACH NES

the tens figure m the second and the n ts figure n the th r l column

The cards are punched by a key p ncl (F1g 61) The card s carr ed by a movi g carr age each depress on of a kev dr ves a



F G 61 HOLLER TH 80 CO U N MECHAN CAL KEY PUNCH

knife through the card and causes the carriage to move one column An expert operator will punch 10 000 or more holes an hour More elaborate punches are ava lable for ta ng ad antage of tl e con stancy of nformat on m certa n co umns of a batch of cards or for reproduc ng the contents of certa n columns from one batch of cards on to columns (not necessarily correspond ng columns or m the same order) of another batch of cards

Th s m t al or punching stage being pract cally the only one into which the human element enters must be carefully checked Th s 1s frequently done by pass ng the cards through a so called

verifier, which is very similar in appearance to the punch, but, instead of knives, has plungers that can pass through holes already made but will not make new holes. The operator repeats the work of the puncher; if a discrepancy occurs, the plunger cannot pass through the card, so the motion of the carriage is arrested, and the error detected. Another method of checking is to print the contents of the cards with the tabulator (see below) and proof-read the list.

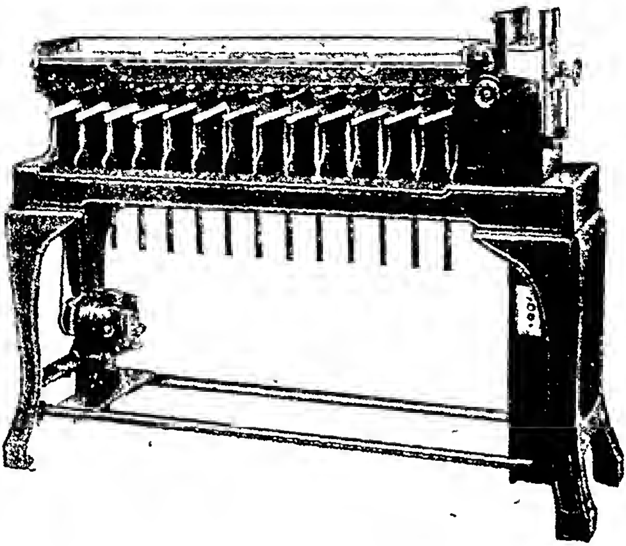


FIG 62. HOLLERITH SORTER

This is perhaps more costly, but is to be preferred in important scientific work, particularly as it enables the final responsibility to be placed with experienced readers rather than with punch operators, who are usually young.

The sorting machine (Fig. 62) examines the cards on any selected column, and distributes them into pockets according to the holes punched in that column. By a systematic sorting over all the columns of any particular field, the cards may be arranged in the numerical sequence indicated by that field. This machine works at the astonishing speed of 24,000 cards an hour. For census and other statistical work, counters are provided to count the number of cards falling into each pocket; in the more advanced forms the numbers obtained in these counters may be printed directly.

multiplication is done nor is any individual product formed only the sums of the products are yielded by the addition of all the partial products In a published description of this method it is related how the sums of nearly a million products were formed in two months—in effect one product per working second The recent *establishment of service bureaux equipped with these machines* has placed them at the disposal of all who may benefit by their use

The Hollerith Multiplying Punch

The Hollerith multiplying punch is a machine that receives its information from punched cards and punches the results of its processes in unused columns on the same card and also accumulates them in a summary counter From two numbers A and B the machine will form AB $A + B$ or $A - B$ From three numbers A B and C it will form various combinations such as $A + B \pm C$ $A + BC$ etc Multiplication is done by a direct process with the result that this is the fastest multiplying machine available Two eight figure numbers may be multiplied and the product punched every five seconds The product thus punched may be used as a later multiplier or multiplicand a feature that is applied for instance when forming the normal equations for fitting a power series by least squares This machine is at present so costly that it cannot be used for ordinary casual multiplications but when the same multipliers or multiplicands are required on different occasions or when the product is required on cards for subsequent sorting or tabulation its use often becomes economical

Adding and listing machines leave a printed record of the items entered into them and of their results Multiplying machines fail to do this although one machine that can print multiplicand and product has been mentioned The Hollerith multiplying punch has the great virtue of leaving a permanent record—albeit in the form of holes—of two factors and their product These holes can be translated into printed figures by the tabulator although this adds to the cost It is highly desirable that this or some other machine should be developed so that it makes at the same time as the multiplication a printed record of some indicative matter the two factors and the product and also provides means for the accumulation of the products and the printing of their total Since such a machine is likely to be expensive it becomes necessary to

ensure that its periods of idleness are as brief as possible; a punched-card feed enables the machine to work continuously, since it may easily deal with the output of several punch operators. Perhaps the machines of the future will use miniature cards, with tiny holes through which rays of light actuate photo-electric cells, so that an entire multiplication may be made and recorded in a second!

REFERENCES

There is no textbook on calculating machines, and the available literature is very scanty. The following papers give descriptions of various machines and scientific applications.

1. "Computing by Calculating Machines." *The Accountants' Journal*, 45, 42 (May, 1927).

Many of the principles of this article remain, although their application to the existing machines has been superseded.

2. "Recent Developments in Calculating Machines." *Office Machinery Users' Association Transactions*, 1927–28.

Although "recent" in 1928, many of these developments are already historical

3. "On the Application of the Brunsviga-Dupla Calculating Machine to Double Summation with Finite Differences." *Monthly Notices of the Royal Astronomical Society*, 88, 447 (March, 1928).

This was the first published account of the application of a commercial calculating machine to mechanical integration. The Brunsviga-Dupla has since been superseded by the Brunsviga 20, the Burroughs (see 4), and the National (see 10)

4. "The Nautical Almanac Office Burroughs Machine." *Monthly Notices of the Royal Astronomical Society*, 92, 523 (April, 1932).

This gives a detailed description of this accounting machine, and of its application to integration from second finite differences and to subtabulation to tenths with third differences negligible. The machine is now superseded, for subtabulation purposes, by the National (see 10).

5. "The Application of the Hollerith Tabulating Machine to Brown's Tables of the Moon." *Monthly Notices of the Royal Astronomical Society*, 92, 694 (May, 1932).

A fuller description of this machine is given in 7. The problem dealt with is the synthesis of harmonic terms.

6. "Computing the Nautical Almanac." *Nautical Magazine*, July, 1933.

A survey of the mechanical computing methods used in the Nautical Almanac Office.

7. "The Hollerith and Powers Tabulating Machines." Printed for private circulation, 1933.

45 56 22 34 48 47 53 52 20 10
14 22 38 37 0 4 50 48 0 34
0 34 (82 examinees)

Tabulate the results in the form of a Frequency Distribution grouping by intervals of 10 marks

8 Plot the results of the above in the form (i) of a Histogram (ii) of a Frequency Polygon

9 From the above data prepare a cumulative table and plot the results

10 Array the data and prepare a rough diagram illustrating the array

11 Reduce the two distributions tabled in Question 24 to a per mille basis and then compare them by means of (i) frequency polygons (ii) ogives

12 Find the geometric means of rows (1) (2) (3) and (4) of Question 7 Compare them with the arithmetic means

13 Find the modes of the distributions in Questions 21 and 24

DERIVATIVE SERIES

14 Check the figures in column (3) of Table 87 using the data in Table 86 Explain briefly what precautions are requisite in order to ensure that the percentages are reliable

15 The International Conference of Labour Statisticians (1923) recommended the following bases for calculating accident rates—

(i) *Accident frequency rate* = number of lost time accidents during period ÷ number of exposure hours (in millions)

(ii) *Accident severity rate* = number of days lost by lost time accidents during period ÷ number of exposure hours (in thousands)

Comment upon these two methods

16 Given the Census population of the United Kingdom 1921 and 1931 explain how you would calculate average imports of foodstuffs per head per annum over the period Criticize the adequacy of the result

17 Crude birth and marriage rates are calculated as so much per 1 000 of the population Criticize this method Can you suggest a better?

AVERAGES ETC

18 From Table 5 (page 40) find the average annual world production of steel (i) for each quinquennium (ii) for each decennium (iii) for the whole period Check the results by averaging the averages

19 Using the same Table verify for the period 1900–1909 that exactly the same results are obtained by averaging—

(i) The excess over 25 million tons
(ii) The defect from 60 million tons
(iii) The differences of the figures from 40 million tons

20 Verify the calculation appearing on page 42

21 The following data relate to sizes of shoes sold at a store during a given year Find the average size (i) by the long method (ii) by the short cut method and verify that the results agree

Size of Shoe (1)	No of Pairs Sold (2)
$4\frac{1}{2}$	1
5	2
$5\frac{1}{2}$	4
6	5
$6\frac{1}{2}$	15
7	30
$7\frac{1}{2}$	60
8	95
$8\frac{1}{2}$	82
9	75
$9\frac{1}{2}$	44
10	25
$10\frac{1}{2}$	15
11	4
$11\frac{1}{2}$	3
12	1
	461

22. Now multiply column (2) (i) by the factor 6; and (ii) by a factor calculated to bring the whole distribution upon a per mille basis. Verify that the averages of the new distributions agree with the figure previously obtained and give an algebraic proof of the theorem.

23. Express your result to the nearest half size. What is there in the nature of this series that makes this process specially appropriate?

24. Find the arithmetic averages of the following frequency distributions by the short-cut method.

Size of Items in feet	and not exceeding	Frequency I	Frequency II
1	10	9	50
10	19	13	70
19	28	86	203
28	37	239	403
37	46	120	304
46	55	46	42
55	64	12	5
		525	1,077

Use 32·5 as the trial average and check by using 23·5.

25. Find the nine-year moving average of the figures of World Production of Steel 1900–29, given on page 40. Assuming this represents the trend, find the annual deviations from trend. Plot the results.

26 Find the weighted average earnings of the female workpeople shown in the table below for 1924 and 1928 taking the numbers to these nearest 1 000 and the earnings to the nearest shilling

Industry	1924		1928	
	No of Workpeople covered	Average earnings	No of Workpeople covered	Average earnings
		s d		s d
Cotton	144 272	29 2	109 848	29 1
Woollen and worsted	86 035	30 11	65 827	30 3
Silk	18 533	27 10	19 700	27 7
Linen	36 722	22 10	25 349	20 6
Hosiery	42 632	28 10	32 085	30 10
Bleaching printing dyeing and finishing	14 797	27 5	12 458	26 4

MEASURES OF DISPERSION, ETC

27 Find the mean mean deviation and standard deviation of the examination marks in Question 7 direct from the individual items

28 Find the same measures from the grouped table resulting from Question 7 Compare the results Would you expect them to differ If so why?

29 Find the mean mean deviation and standard deviation of the sizes of shoes tabled in Question 21

30 Find the mean mean deviation and standard deviation of the sizes of items tabled in Question 24 Then compare their means and coefficients of dispersion

31 Find the median and quartiles of the examination marks given in Question 7 (i) by the direct method (ii) by simple interpolation from the grouped date and (iii) by a graphic method Would you expect the results to differ If so why?

32 Find the median and quintiles of the distribution of sizes of shoes given in Question 21 by the direct method

33 Compare the distributions of Question 24 by reference to their first and ninth deciles

34 Find the quartile deviations of the distributions of Question 24 and then find the corresponding coefficients

ERROR AND APPROXIMATION

35 Re write the figures of total population of the United Kingdom from the *Statistical Abstract for the United Kingdom* correct (i) to the nearest 10 000 (ii) to the nearest million What is the amount of error involved? What type of error is it?

36 Assume you were given the male and female population of England and Wales 1881–1931 correct to the nearest 100 000 Write

down the ratios of males to females taking the male population as unity and indicate the precision of your results.

37. In Table 77 (page 226) assume the weights are subject to possible errors of 0·3, 0·3, 0·2, 0·2, and 0·1 respectively, while the percentage increases are subject to possible errors of 2, 3, 4, 1, and 1 points respectively. What is the possible error in the result?

38. Find the values of—

(i) $(82 \pm 4) + (59 \pm 6)$.
(ii) $(463 \pm 10) - (21 \pm 5)$.
(iii) $(700 \pm 35) \times (60 \pm 6)$.
(iv) $(500 \pm 25) \div (30 \pm 3)$.

39. A factory turns out an article by mass production methods. From past experience it appears that 10 articles on the average are rejected out of every batch of 100. Find the standard deviation of the number of rejects in a batch. What is the maximum number of rejects per batch likely to be encountered?

40. It is reported that several batches have recently been turned out containing 20 to 30 rejects. What inference would you draw?

41. Give formulae for the error of a total of n estimates upon the assumption (i) that the individual errors are cumulative. (ii) that they are compensating. What presupposition is involved by formula (ii)?

INTERPOLATION.

42 The following table shows the value of an immediate life annuity for every £100 paid—

Age (years) . .	40	50	60	70
Annuity(£) . .	6·2	7·2	9·1	12·8

Prepare a difference table and interpolate for ages 42, 54. 57, and 69.

43. Using a four place table, find the logarithms of 20, 30, 40, 50, and 60. Then find the logarithms of 25, 33, 46, and 57 by interpolation. Check the results from the table.

44. The following figures are taken from the sur-tax statistics 1928–29—

Income	*No. of Persons*
£2,000–£3,000	42,737
£3,000–£4,000	20,262
£4,000–£5,000	11,229

Estimate the number of persons between £2,500 and £3,000 by interpolation. The actual number was 17,337. Why do the two results differ?

45. The following figures relate to the number of estates liable to Estate Duty, 1930–31—

Class of Estate	*No. Liable*
£25,000–£30,000	638
£30,000–£40,000	740
£40,000–£50,000	415

Estimate the number between £31,000 and £32,000 by interpolation

46 Fit straight lines to the following series by the method of factorial moments and plot the results

Year—	1	2	3	4	5	6	7	8
Series (i)—	810	842	890	801	852	899	871	919
Year—	9	10	11	12	13			
Series (i)—	921	890	985	1182	1564			
Year—	1	2	3	4	5	6	7	8
Series (ii)—	1871	1866	1881	1137	1212	1340	1287	1394
Year—	9	10	11	12	13			
Series (ii)—	1302	1278	1328	1267	1073			

47 From the following data (drawn from biological sources) find the correlation coefficients

(i)	$X =$ 39	40	41	42	43	44	45	46	47	48	49	50	51
	$Y =$ 70	70	71	72	73	73	74	75	75	77	77	78	79
	$X =$ 52	53	54	55	56	57	58	59	60	61			
	$Y =$ 79	80	80	81	82	82	82	83	85	88			
(ii)	$X =$ 68	69	70	71	72	73	74	75	76	77	78	79	80
	$Y =$ 40	41	41	14	44	45	46	47	49	49	50	51	52
	$X =$ 81	82	83	84	85	86							
	$Y =$ 53	54	55	57	56	58							

INDEX NUMBERS

48 Using the seven series of price relatives in the lower part of Table 44 calculate the following—

(i) A weighted index number with weights
10 5 3 1 1 10 10 respectively

(ii) Ditto weighting the coal and iron items 10 each and the rest 1 each

(iii) An index based on the unweighted geometric mean

(iv) A chain base index based on the arithmetic mean

49 Give a formula for testing the reliability of an index number assuming all the items were independent Would it apply if they were not independent?

MISCELLANEOUS QUESTIONS

50 How would you adjust a sterling series for changes in the purchasing power of the £?

51 Country A s records of exports to country B do not in general agree with country B s records of imports from country A Give reasons

52 The average age of His Majesty s judges is greater than that of professional cricketers What inferences do you draw?

53 It is required to estimate the population of a large town The ast Census was taken in 1931 What means do you suggest?

ANSWERS TO EXERCISES

(7) 7, 4, 9, 15, 7, 16, 10, 6, 4, 4. Total = 82.

(12) G.M. 63·56, 55·44, 31·45, 63·14.
A.M. 66·90, 59·20, 38·60, 65·50.

(13) Size 8, 33·06 ft. (33·24 ft. by customary formula); 34·02 ft. (33·40 ft. by customary formula).

(21) Size 8·40.

(24) 34·37 ft., 31·75 ft.

(27) 46·24 marks, 19·91 marks, 24·22 marks.

(28) 46·57 marks, 19·93 marks, 23·93 marks.

(29) Size 8·40, 0·85, 1·10.

(30) (I) 34·37 ft., 7·24 ft., 9·69 ft.
(II) 31·75 ft., 7·79 ft., 10·42 ft.

(31) (i) 48 marks, 30 marks, 64 marks.
(ii) 48·1 marks, 29·8 marks, 63·0 marks.

(32) Size $8\frac{1}{2}$, $7\frac{1}{2}$, 8, $8\frac{1}{2}$, $9\frac{1}{2}$.

(33) (I) 22·2 ft., 47·1 ft.
(II) 17·4 ft., 44·2 ft.

(34) 5·86 ft., 6·79 ft., 0·17, 0·21.

(39) 3, 19.

(42) £6·37, £7·80, £8·38, £12·32.

(44) 17,719 persons.

(45) 94 estates.

(46) (i) $Y = 685{\cdot}22 + 38{\cdot}62\,X$.
(ii) $Y = 1761{\cdot}62 - 51{\cdot}24\,X$.

(47) + 0·9887, + 0·9958.

INDEX